OCTOBER
1954
VOLUME 18
NUMBER 5

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

Edited by Laurance F. Shaffer

Teachers College, Columbia University

Associate Editors: Edgar A. Doll, *Bellingham, Wash.* Ann Magaret Garner, *University of Illinois.* William A. Hunt, *Northwestern University.* E. Lowell Kelly, *University of Michigan,* Morris Krugman, *Board of Education, New York, N. Y.* Berina M. Luckey, *Board of Education, Cleveland, Ohio.* Fred McKinney, *University of Missouri.* Anne Roe, *New York, N. Y.* Carl R. Rogers, *University of Chicago.* R. Nevitt Sanford, *University of California.* Edward Joseph Shoben, Jr., *Teachers College, Columbia University,* J. R. Wittenborn, *Rutgers University.*

Lorraine Bouthilet, *Managing Editor*

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

Editorial Office

Manuscripts submitted for publication and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y. All manuscripts should be submitted *in duplicate* to facilitate referral to the Associate Editors. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

Style of manuscripts. Manuscripts should be prepared in the style described in the Publication Manual of the American Psychological Association (*Psychological Bulletin,* 1952, 49, 389-449). The Manual may be purchased from the Washington office of the Association for $1.00. Otherwise acceptable manuscripts which depart from the standards of the Manual are returned for revision and retyping.

Brief Reports. A one-page condensation of an acceptable research report may be given early publication without charge if the author agrees (*a*) to make 100 mimeographed copies of the full report available to those who request them, (*b*) to send a copy of the full report to the Editor for deposit in the American Documentation Institute, and (*c*) not to have the full report published elsewhere. For detailed instructions on how to prepare a Brief Report, see the back of the *Contents* page of any 1954 issue of this Journal or write to the Editor.

Early Publication. Authors may secure the early publication of an acceptable article in the first issue to go to press after its acceptance by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. The charge to the author for such early publication is $20.00 per printed page.

Reprints. Authors receive fifty reprints gratis of all articles except early publication articles and Brief Reports. Additional quantities may be ordered when returning proof.

Business Office

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St. N. W., Washington 6, D. C. Address changes must reach the subscription office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

The *Journal of Consulting Psychology* is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 480 pages. The subscription per year is $7.00, foreign $7.50, single number $1.50.

Copyright, 1954 by the American Psychological Association, Inc. Printed and mailed by the Dentan Printing Co., 23 West Colorado Ave., Colorado Springs, Colo. and entered as second-class matter at the Post Office at Colorado Springs, Colo. Acceptance for mailing at special rate of postage provided for in section 538, Act of February 25, 1925, authorized May 27, 1940.

JOURNAL OF CONSULTING PSYCHOLOGY

October, 1954 Vol. 18, No. 5

CONTENTS

An Assessment of the Current Role of Psychodiagnostic Testing: *Alan K. Rosenwald* - - - - - - - - - - - - - 311
Acceptance and Acceptability: An Examination of Social Reciprocity: *Seymour L. Zelen* - - - - - - - - - - - - - - - - 316
Rorschach Color Responses and Manifest Anxiety in College Women: *Wayne H. Holtzman, Ira Iscoe, and A. D. Calvin* - - - - 317
Motor and Fantasy Correlates of Rorschach Human Movement Responses: *Jerome L. Singer and Jack Herman* - - - - - - - - - - - - - 325
The Ammons FRPV Test and the WISC for Remedial Reading Cases: *Louis M. Smith and Arline R. Fillmore* - - - - - - - - - - - - 332
A *P* Scale for the Rorschach: A Methodological Study: *Philip H. DuBois and Arnold H. Hilden* - - - - - - - - - - - - 333
Sex, Age, and Iconicity as Factors in Projective Film Tests: *Charles J. McIntyre* - - - - - - - - - - - - - - - - - - 337
A Note on the Use of a Paper-and-Pencil Form of the MMPI *Hs* Scale for Hospital Use: *Sol Charen* - - - - - - - - - - - - 344
Transcendence Indices, Extent of Personality Factors in Fantasy Responses, and the Ambiguity of TAT Cards: *Douglas T. Kenny* - - - - - - - - - - - 345
A Psychotherapy Case Showing Progress but No Decrease in the Discomfort-Relief Quotient: *Edward J. Murray, Frank Auld, Jr., and Alice M. White* - - 349
The Effect of Childhood Diseases on IQ Variability: *Lillian Wagenheim* - - - - - 354
The Relationship Between Motor and Cognitive Inhibition: *Julian Meltzoff and Murray Levine* - - - - - - - - - - - - 355
Classification of "Suicidal" Patients: *Albert Rosen, William M. Hales, and Werner Simon* - - - - - - - - - - - - - - - - - - 359
Duration of Illness as a Prognostic Indicator in Mental Illness: *C. H. Swensen, Jr. and G. R. Pascal* - - - - - - - - - - - - 363
Rigidity in Conceptual Thinking in Exogenous and Endogenous Mentally Retarded Children: *J. G. McMurray* - - - - - - - - - - - - - 366
Test Performance of Preschool Children With and Without Organic Brain Pathology: *Miriam Stearns Haines* - - - - - - - - - - - 371
Sex Differences on the Revised Beta Examination: *George Lavos* - - - - - - - - 375
A Validation of a Bender-Gestalt Scoring System: *Murray Lonstein* - - - - - - - 377
Gellerman's Complex Crossover Design: *Julian C. Stanley* - - - - - - - - - - 380
Mental Age Equivalents for the Revised Stanford-Binet Vocabulary Test: *Edward E. Cureton* - - - - - - - - - - - - - - - 381
New Books and Tests - 384

Brief Reports:

A New Publication Procedure

For a trial period, the *Journal of Consulting Psychology* will accept *"brief reports"* for early publication without expense to the author, under certain conditions. The new procedure is intended to reduce publication lag and permit the circulation of worth-while studies of specialized interest or minor importance which cannot be accepted at present because of limitations of space.

A *brief report,* strictly limited to one printed page, may be accepted if the author agrees to make a more extended report available. Up to four pages in each issue may be devoted to brief reports, published in the order of their receipt without regard to the dates of receipt of articles accepted for regular publication. It is anticipated that most of the brief reports can appear in the first or second issue to go to press following their acceptance.

The procedure for submitting a *brief report* is:

1. The author sends the editor *both* the brief report and a more complete report of the study. The full report should be of sufficient length to give a clear account of the background, procedure, results, and conclusions of the study.

2. The *full report* is prepared in the style specified by the *Publication Manual* [1], except that it may be typed with single spacing. The editor will send the full report to the ADI.

3. The *brief report* gives a clear, condensed summary of the procedure of the study, and as full an account of the results as space permits. To insure that the brief report does not exceed one printed page, the author prepares it according to these specifications:

The text of the brief report, including all matter except the title and the author's lines, does not exceed 80 typewritten lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, or 4.2 inches long in pica. The manuscript is double spaced throughout and, except for the short lines, follows the standard style [1].

Headings, tables, and references are avoided in the brief report. If essential, they must be counted in the 80 lines. The brief report includes, within its 80 lines, a footnote in this style:[1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. [giving the author's full name and address] or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. .... from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $..... for microfilm or $..... for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

4. The author of a brief report prepares at least 100 mimeographed copies of the extended report. He agrees: (*a*) to send copies of the extended report upon request as long as the supply lasts, and (*b*) not to submit the extended report for publication to another printed journal of general circulation.

Reference

1. American Psychological Association. Council of Editors. Publication manual of the American Psychological Association. *Psychol. Bull.,* 1952, 49, 389-449.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

An Assessment of the Current Role of Psychodiagnostic Testing[1]

Alan K. Rosenwald
University of Illinois, College of Medicine

In any attempt to consider the current role of psychodiagnostic testing in clinical psychology, one can see analogues in the divergent opinions and shades of thought characteristic of our political scene.

Our various media of current opinion have given much, perhaps too much, consideration to the varied political philosophies adopted by this or that person. Liberal, conservative, nationalist, internationalist, leftist, rightist, etc. have become common terms of approval or opprobrium. To a lesser degree, it seems that psychologists are just as likely to stereotype their fellow psychologists by some equally superficial characteristic. The clinician decries the experimentalist as lacking in certain qualities, chiefly human; the experimentalist condemns the clinician by use of the telling, if overly used, phrase, "He's no scientist." Clinicians war with other clinicians so that "the battle lines are drawn." Projectivists clashing with the nonprojectivists, objectivists with the intuitionists, statisticians with the qualitative analysts, and Freudians with the non-Freudians are illustrative of the strife in which each and everyone of us becomes involved.

In this evaluation of the current major trends in diagnostic testing and in trying to anticipate our hopes or expectations of the future, the writer, too, finds himself forced into the use of labels. With this injected note of warning about possible oversimplification, the writer would wish to consider the trend, or label, illustrated by the school of thought he has designated as conservative. The conservative seeks to demonstrate the validity of his diagnostic instruments. In fact, the concern seems more than primary; it seems to be an all-encompassing one. This evaluative statement should not be construed as an attempt to decry validity and/or reliability studies, nor is it a passing glance at superficially recognizing the need for such studies only to embark on a series of psychological speculations which lack any scientific basing. It is a matter of concern, however, that the literature abounds in numerous studies which by their design seem almost destined to purposefully devalue certain tests.

As an example of the validational or so-called scientific approach, a study by Wittenborn illustrates the grievous principle when he says, "One of the outstanding characteristics of the Rorschach responses is their highly individual quality, their extreme variability from person to person. This makes conventional statistical treatment of Rorschach data awkward because of the difficulty in securing a large number of roughly identical responses. For example, out of a group of one hundred normal subjects many, if not all, would report human movement responses but relatively few would have more than two, or at the most, three identical responses in common. In the present study an attempt has been made to reduce this difficulty by the use of the Harrower-Erickson checklist" [6, p. 258]. Wittenborn then notes that some question has been raised as to the validity of this check list for diagnostic purposes, but defends his use in this particular study since he wants to investigate the interrelationship of various scoring factors, specifically, movement and color. It would seem that such a defense is, at best, questionable, since if the check list is invalid for diagnostic purposes, then one would logically question whether or not the responses gained from such a procedure could permit one to make sound inferences about the Rorschach method in its clinical setting. Moreover, it would seem that

[1]Presented as part of the symposium, "Assessment of Current Trends in Psychodiagnosis," at the APA meeting, Cleveland, September, 1953.

there is a potential danger to the reasoning which Wittenborn uses. According to his reasoning, if we have a particular test which fails to lend itself to being treated by orthodox statistical procedures, then the test itself must be altered so as to render it suitable for statistical manipulation. The reasoning is something like this: a switch to an invalid check list permits the data to be processed statistically. In other words, if a particular phenomenon can't be measured, measure an artifact of the phenomenon. On the basis of data gained from measuring the artifact, one can then turn to generalizing about the original phenomenon.

As illustrative of the dynamicist or the "radical" approach, one can turn to Gurvitz's volume, *The Dynamics of Psychological Testing* [3]. In a particular instance, he is evaluating the association a patient made to a female figure drawing. When asked for the worst part of the drawing, the patient replied "Toe nails; they're ingrown, that's why she's not wearing shoes." Later on, in his analysis of the test protocol, Gurvitz says, "The ingrown toe nails appear to represent the self-punishing and masochistic elements in the personality" [3, pp. 116-119]. The existence of masochistic and self-punishing elements in a personality as stated is of no moment. Actually it would be a far more unusual statement in a psychological report if one could read that patients' masochistic and self-punitive elements were of little import in their personality. Certainly it is difficult to rule out the possible existence of such elements in personality formation; one can, however, question the basis on which the inference was made. Before reading his conclusion, the writer would have assumed that the patient was possibly attempting to be facetious and, by the same token, critical of the question or questioner. If speculation were the order, one might even say that the patient was trying to rationalize her economic deprivation, i.e., lack of shoes, by focusing on the somatic. At any rate, the elemental importance of the statement doesn't seem so great that it permits the delineation of a major aspect of personality.

Apparently many diagnosticians are doing what some psychoanalysts did a number of years ago (some still do) in attaching undue importance to everything that a person may say. Also, it would seem that we are in danger of covering our failure to understand certain personality operations by trying to couch our reports in high-sounding, esoteric language, which simply serves to make the person so described something less than a living being. One need not be unmindful of the danger that attempts to make reports too individualized may lead us on an unscientific road. If we get too enamored with our own literary creativeness so that we neglect to make certain general statements which should be made about an individual, then an injustice is being perpetrated. When these general statements become "he is an immature individual" or "his problems of hostility have not been solved" or what has now become the most cliché of psychological clichés, "his intellectual efficiency does not measure up to his intellectual potential," then it would appear that our reports have become jaded, lacking in specific statements about an individual, and are becoming a mass of generalities which unfortunately don't even glitter.

Holt's analysis of the TAT of Jay, a case study, illustrates too well the point. Before setting down to analyze the results, Holt gives his ESP interpretation. "As a kind of control, I decided to set down, before looking at the TAT some of the truisms and clichés that are found in most test reports, and that can be fairly safely said once you know someone is a patient of Jay's age and sex" [5, p. 457]. Then follows this ESP interpretation, which makes perfectly good reading and is just as definitive as many psychological reports that thrive in case files throughout the country.

Much of what has been said has concerned itself with the varying attitudes held toward projective tests. Many of the same observations can be made if we examine the area of intelligence testing. The following statements are found in Patterson's manual: "His social judgment is relatively superior, indicating that he has been capable of profiting by his experience. In contrast, however, foresight and planning ability (Picture Arrangement) is poor. This poor performance may have been affected somewhat by his difficulty of attention and concentration which may also have affected the Picture Completion and Digit Symbol as well as the Digit Span tests. Picture Arrangement and the Similarities as well as Vocabulary may also be affected by cultural background. In

the Picture Arrangement test, after successfully arranging the first group of pictures from left to right, he arranged the next group from right to left and could give no explanation when asked to; he immediately placed the next series (the elevator) in a vertical order" [4, pp. 36-37]. Here again, we are confronted with interpretations that fail to account for discordant elements in the data as well as end up by denying the interpretation which has just been given. If we have an individual whose social judgment is relatively superior, then it would seem that his foresight and planning ability should also be superior. If, however, the performance of a patient is affected by both lack of concentration and attention, as well as poor cultural background, then perhaps we are not permitted to infer what an individual's foresight and planning ability are. Maybe we should say that he is so distractible that we could infer he could not plan ahead in a prudent manner. We must question whether we have a measure of a person's social judgment merely because he has a high score on Comprehension. Patterson does imply in the quotation above that his social judgment is an index to what the patient was like prior to his illness and that he had been capable of profiting from past experiences. If that is the case, why mention a poor cultural background, which would certainly connote a lack of experiences, common to all of us, from which he could profit. Further, why is there no emphasis on the confusion which surrounds the performance on the Picture Arrangement test and which would warrant considering it the production of a brain-damaged individual.

These examples from Patterson, Gurvitz, and Wittenborn have been selected at random from the literature to illustrate certain trends in diagnostic testing, rather than as three particular studies with which the writer wishes to quarrel. Certain conclusions should be drawn from these types of studies. First, it seems that our need for purification of our diagnostic instruments has led us all too frequently into neglecting how we get data, but then becoming extremely precise in our measurement of very unprecise primary data. Secondly, our need to demonstrate our "know-how" has caused us to engage in many speculative statements, of which a great number cannot be validated because we have no criteria by which to demonstrate their validity. In other words, their world-wide applicability to man makes our individual patient no different from anybody else in the world. Finally, we apparently seem unable to ever categorically and unequivocally state that we don't know the answer to a particular question — in other words, that we just don't have all the answers to all the questions. Rather we seem to get involved in a series of statements, some of which say one thing, followed by statements which deny what we have just said. Perhaps this is a tribute to our capacity for tolerating ambiguity; it does not, however, make for definitive diagnostic interpretations.

One should pause and consider some of the uses of diagnostic testing in a psychiatric setting. It would lead us far afield to consider the many different kinds of problems upon which we may be asked an opinion. We can say most referrals in a psychiatric clinic seek to know something about a person's intelligence, both with respect to over-all capacity as well as strengths and weaknesses shown in the separate functions, something of the personality structure, some of the attitudinal constellations which are dominant in a person, etc. More broadly than this, however, is the implied question, "What sort of a person is this?" Presuming the person has pathology, what sort of expectation do we hold regarding the course of illness; what can we expect in therapy? The writer feels that most referrents are not primarily interested in how we arrive at conclusions, providing, of course, that we do not damage the patient in the process of reaching our conclusions. They are interested in our conclusions, either because they constitute a check on their subjective impressions or because they reveal areas of personality functioning which they had not been able to explore adequately. There is danger of belaboring the obvious, but it seems well to consider what we are doing before we try to say what we should be doing. If we boil down the major functions of diagnostic testing as aiding in differential diagnosis, saying something about prognosis in or without therapy, and always recognizing that diagnostic instruments are useful research tools, why the need to assess the role of testing? The need to establish its role grows from a need to define its value. This concern is en-

hanced by the declining emphasis on diagnostic testing. In recent years clinical psychologists have shifted overwhelmingly in the direction of becoming therapists — and diagnostic testing has become confined by the "expert" telling the "nonexpert" how to examine and how to interpret results, what to do, what not to do, etc. Rarely does he do what he teaches. Moreover, the "average" person who comes for training in clinical psychology usually has in mind training in therapy or doing research, and only secondarily training in the diagnostic study of patients. The reasons which underlie the subordination of diagnostic testing to therapy and research are too numerous and involved to consider here. At any rate, the result is evident. Diagnostic testing may be likened to the status of a second-class citizen who is tolerated and endured but who is pegged below his superiors.

The writer questions strongly the desirability of this rejection of diagnostic testing for therapy. In turning to therapy, clinical psychologists are neglecting areas of study which could occupy their interest but which do not. This feeling does not stem from any antihumanitarian or antitherapeutic origin. Certainly psychological research in the therapeutic process will lead to some very valuable contributions. The criticism, however, is that there are problems in clinical diagnosis which require careful study, and that these will be neglected as the clinical psychologist focuses on being the practitioner.

Thus, all of us speak very seriously of the need to study the normal personality, yet our words are, for the most part, empty ones. Ask yourself how long it has been, if ever, since you wrote a report in which adequacy of function was the keynote of your report rather than a person's many inadequacies. We are now very proud of our newly acquired dynamic orientation which permits us to look askance at the work of Kraepelin, Bleuler, or any of the more descriptive psychiatrists. Instead, we speak knowingly of weak ego functioning, poor reality testing, latent schizophrenia, etc. until we are caught up in a series of ritualistic statements which, although colorful, fail to help us understand the patient.

We are victims of what Herbert Block, the *Washington Post* cartoonist, in his *The Herblock Book* calls *Mumbo-jumbo.* "One of the things that makes life complicated . . . is the practice of Mumbo-jumbo. Mumbo-jumbo, as I use the term, is the transference of devotion from a set of principles to the symbols and rituals that are built around them. It's the business of taking over a good sound idea, building an elaborate shell around it, then carefully removing the already wilting idea and beating its brains out with the shell. Mumbo-jumbo creeps up in practically every field of endeavor. It gets into the arts . . . into medicine when the operation becomes more important than the patients" [2, pp. 41-42].

Granted that we lose the patient in the process of studying him, why should this be? Partially, it is because we have failed to see the patient as he lives from day to day. In our assessment of people, we are seemingly at a loss in predicting such elementary things as how the person gets along with other persons, whether a patient is liked or disliked by other patients, what sorts of people he likes or dislikes, who the leader of a psychiatric ward might be, how the patient would get along with certain authority figures, etc., etc. It is not unreasonable to propose that validational studies should concern themselves with such problems. The writer concludes, a priori, that there is nothing wrong with psychologists spending time observing the behavior of people. As an example, one can see much merit in the approach of Barker and co-workers [1]. The time-worn story about a patient's remark to a psychiatrist still holds. "What do you know about a patient? You only see him once, the nurses and attendants are here only eight hours a day, but me, I'm here a full 24 hours a day. What do you want to know about so and so?"

Increased knowledge of the ins and outs of a patient's life should lead us to try and cast our studies in a more practical mold. Not only that, but it might lead us to use our diagnostic instruments as a means of assessing the person as he lives from day to day rather than describing him by means of a series of generalized statements which are impressive and high-sounding, but not very meaningful.

In closing, the writer would like to re-emphasize his feeling that diagnostic testing should further our understanding of personality structure and add to our limited knowledge

of psychopathology. If we glance back at what has been written, much has been couched negatively. This negativism, as it were, has its justification in a feeling that we are not pressing toward the solution of major problems in clinical psychology which could be aided by diagnostic testing. Our literature is swollen with papers reporting the administration of good clinical programs, validation or invalidation of countless tests, and innumerable brochures and manuals on interpretation of this or that test. We seem to be avoiding any long-term problems in clinical psychology; administrative planning and the mushrooming development of new tests with all the standardization problems lend our work a certain frenetic quality. Is it an idle hope that we will move in the direction of long-term studies which seek to answer certain basic problems.

Received February 22, 1954.

References

1. Barker, R. G., & Wright, H. F. *One boy's day.* New York: Harper, 1951.
2. Block, H. *The Herblock book.* Boston: Beacon, 1952.
3. Gurvitz, M. S. *The dynamics of psychological testing.* New York: Grune & Stratton, 1951.
4. Patterson, C. H. *The Wechsler-Bellevue scales, a guide for counselors.* Springfield, Ill.: Charles C. Thomas, 1953.
5. Shneidman, E. S., Bell, J. E., Fine, R., Holt, R. R., Kutash, S. B., & Vorhaus, Pauline G. The case of Jay: interpretations and discussion. *J. proj. Tech.*, 1952, **16**, 444-475.
6. Wittenborn, J. R. Statistical tests of certain Rorschach assumptions: analyses of discrete responses. *J. consult. Psychol.*, 1949, **13**, 257-267.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Acceptance and Acceptability: An Examination of Social Reciprocity[1, 2]

Seymour L. Zelen
State University of Iowa

One hundred and forty-five sixth-grade children in five classes were given the Bonney Sociometric, the Feelings of Personal Worth Subscale of the California Test of Personality, and the W-A-Y technique. The latter two instruments yielded scores of self-acceptance, while the sociometric gave scores of peer status and acceptance of peers.

Peer acceptance correlated .59 with acceptance of others, significant at the .01 level, and .30 and .39 with the self-acceptance scores of the W-A-Y and the CTP Feelings of Personal Worth Scale respectively, both significant at the .01 level. Acceptance of others correlated .08 with the W-A-Y and .10 with the Personal Worth Scale; neither of these proved significant.

The expected positive correlations of peer acceptance with acceptance of others and acceptance of self were obtained. This suggests that liking others and being liked by others are reciprocal characteristics. Individuals with high group status tend to find many group members highly acceptable. Self-perception and perception by one's peers also produced a significant relationship. These findings suggest that a child who has positive feelings about himself is better able to devote his energies to the group activities and to cooperate more fully with others. The child with negative self-percepts may be constantly on guard against new perceived threats from others.

The lack of any relationship between the measures of self-acceptance and acceptance of others was unexpected. An interpretation is offered in terms of the discriminability of interpersonal or social cues.

It is possible that children might be aware of and capable of discriminating cues of external behavior, but would not be aware of the cues upon which the more subtle social relationships are based.

Brief Report
Received June 24, 1954.

[1]An extended report of this study may be obtained without charge from Seymour L. Zelen, Child Welfare Research Station, University of Iowa, Iowa City, Iowa, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4307 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2]This research was supported by Research Grant MH-301 from the National Institute of Mental Health, of the National Institutes of Health, U. S. Public Health Service.

References

1. Bonney, M. E. Personality traits of socially successful and socially unsuccessful children. *J. educ. Psychol.*, 1943, **34**, 449-472.
2. Bugental, J. F. T., & Zelen, S. L. Investigations into the 'self-concept'. I. The W-A-Y technique. *J. Pers.*, 1950, **18**, 483-493.
3. Clark, W., Tiegs, E., & Thorpe, L. California test of personality: Form AA, Intermediate. Los Angeles: California Test Bureau, 1953.
4. Sheerer, E. An analysis of the relationship between acceptance of and respect for the self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13**, 169-175.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Rorschach Color Responses and Manifest Anxiety in College Women

Wayne H. Holtzman, Ira Iscoe
The University of Texas
and A. D. Calvin[1]
Michigan State College

In recent years there has been a sharp increase in the number of experimental studies concerning the validity of the Rorschach test. In particular, a group of investigations has appeared dealing with the nature of reactions to color. In reviewing these studies, Benton [2] and Hertz [9] point out the generally negative findings with respect to the effects of color and emphasize the need for considerable revision of earlier Rorschach theory. Most of these studies are limited in that they dealt with the validation of only one or two Rorschach scores or considered only one of many pertinent independent variables.

The present research was designed to investigate the effects of anxiety, color, stimulus configuration, order of presentation, and examiner, considered singly and in combination, upon a large number of Rorschach variables.

Description of the Study

The Minnesota Multiphasic Personality Inventory (MMPI) was administered to 696 college women in various dormitories and sororities at The University of Texas. Answer sheets were scored for both the Taylor Manifest Anxiety Scale [17] and the Winne Scale of Neuroticism [20]. High- and low-anxiety groups were established by selecting women having both Taylor scores in the upper or lower 15 per cent and Winne scores in the upper or lower 25 per cent of the respective distributions. Greater emphasis was placed upon Taylor's scale than Winne's because of its higher reliability and its more extensive use in previous experimental studies of anxiety [10]. One hundred and twenty-eight women, 64 from each criterion group, served as subjects in the experiment.

Achromatic copies of Rorschach Cards II, III, VIII, and IX were carefully prepared by a special photographic process to insure faithful reproduction of the inkblots. These particular cards were chosen for the following reasons. First, Cards II and III represent one kind of color application while cards VIII and IX represent another. The former have but a single color, bright red, while the latter have a variety of colors blending together. It is quite possible that one type of color would have a different effect than the other. Second, Cards II and III are only partly colored and partly black and white, whereas Cards VIII and IX are completely colored. The advantages of using these two pairs of cards rather than any others will become apparent when the experimental design has been discussed.

Since it was desirable to study simultaneously a number of independent variables in order to isolate interactions as well as main effects, a rather elaborate experimental design was employed. Thirty-two basic plans were constructed in which stimulus configuration or inkblot, color, and order of presentation were varied systematically in a completely balanced fashion. One half of the subjects received a colored card on the first trial, the other half getting an achromatic card first. Similarly, each stimulus configuration appeared equally often on each trial. A standard procedure with full inquiry, according to Klopfer and Kelley [11], was used; each individual was administered all four inkblots, two in the original Rorschach version and two in the achromatic form.

The number of possible permutations or

[1]Formerly at The University of Texas.

orders of presentation of the four inkblots, ignoring the presence or absence of color, is 24. This was reduced to eight by pairing Card II with Card III and Card VIII with Card IX. Thus, both members of a given pair always appeared together in either the first or last two trials. When the color variable (chromatic vs. achromatic) was superimposed on stimulus configuration, a total of 32 plans was required to balance the design completely.

With 128 subjects, four replications of the basic design were possible. Two additional independent variables, Examiner X vs. Examiner Y, and high-anxiety group vs. low-anxiety group, used up the replications. Since much of the confusion in the Rorschach literature may be a function of examiner variance, two examiners were used, one with considerable experience in administering and interpreting the Rorschach and the other with formal training on the Rorschach but little clinical experience. Subjects within each anxiety group were assigned at random to examiner and plan by a third person, the only individual who knew whether a particular subject belonged to the high- or low-anxiety group. This was necessary to avoid the possibility that the examiner's administration of the cards might be influenced by knowledge of the subject's classification.

Every Rorschach response was scored for location, form level, movement, shading, and content. Because of the subjectivity necessarily involved in scoring form level, each response was scored twice according to Beck's tables for $F + \%$ [1] by two psychologists working independently. The interscorer reliability for $F + \%$ was .83. The actual scoring reliability was probably higher since every scoring discrepancy was resolved by a conference of all three authors. The number of responses, the reaction time, and any overt remarks by the subject were also recorded for each card. In addition, every protocol was scored according to Elizur's signs of anxiety [8], a list of signs based on content analysis.

Results

Since all possible advance clues to the subject's classification were withheld from the examiners, the more experienced of the two examiners (X) judged each of his 64 subjects as "anxious" or "nonanxious" solely on the basis of behavior during the individual testing session. No judgments were made by the other examiner since he did not feel sufficiently competent to attempt this task. A number of signs were employed by X in making a judgment: the degree of composure shown by the subject upon entering the room, questions regarding the purpose of the test, the subject's attitude toward X, his readiness to cooperate, and his attentiveness to directions. After administering the test, if X still felt that he needed more information before making a final judgment, he conducted a short interview in which attitudes toward the university, school grades, social activities, and plans after graduation were discussed. The basis of judgment could best be described as intuitive and global, consisting primarily of behavioral signs other than specific responses to the test materials. A comparison of the judgments made by X with the actual criterion grouping is presented in Table 1. That he was able to detect the high- and low-anxiety groups with such a high degree of accuracy attests to the validity of the anxiety criterion.

Table 1

Success of Examiner X in Judging Anxiety Solely on the Basis of Subject's Behavior during the Testing Period

Examiner's judgment	Grouping based on Taylor's scale		
	High anxiety	Low anxiety	Total
High	26	9	35
Low	6	23	29
Total	32	32	64

$\chi^2 = 16.14$ ($p < .001$).

Three general levels of analysis were made: one based upon response to the first card only in the series of four cards presented to each subject; one dealing with only the first two cards in the series; and one concerned with data such as total number of responses and $F + \%$ arising from all four cards.

When only the first card is considered, four of the primary variables—examiner, anxiety, color, and Card II vs. Card III—can be arranged in a 2^4 factorial design with four replications to provide an adequate error term. A similar design exists for Cards VIII and IX.

Ignoring the differences cited earlier between these two pairs of cards would allow combination of both designs with double the number of cases for evaluating the effects of examiner, anxiety, color and their interactions.

The dependent variables or test scores that were analyzed according to this design consisted of the following: (*a*) reaction time, the time from initial exposure of the card until the first response; (*b*) number of responses; (*c*) number of responses using the whole inkblot; (*d*) $F + \%$; (*e*) form level of the first response only; (*f*) number of responses having some kind of movement as a determinant; and (*g*) number of responses with shading as a determinant.

Two general kinds of statistical procedure were used throughout the study. When the test scores were fairly continuous and normal (or could be readily transformed into a normal distribution), analysis of variance methods were employed. At the first level of analysis this proved to be the case only for reaction time. When the test scores were categorical in nature or relatively discrete and severely skewed, the distributions were dichotomized near the median and handled by chi square or Fisher's exact method [6] for testing the significance of the difference between two percentages.

The second level of analysis dealt with the first two trials together rather than Trial 1 alone. The use of a two-by-two latin square combined with a factorial design made possible an analysis of the effect of order of presentation of the color (colored-before-achromatic vs. achromatic-before-colored) as well as practice effects, anxiety, color, and the interactions of anxiety with color, trial, and order. This was done separately for Cards II and III and for Cards VIII and IX. Differences among the stimulus configurations were completely lost owing to confounding. This design was employed for the analysis of reaction time only since the other variables were too discrete and skewed.

The third level of analysis was concerned with data obtained from the subject's performance on all four cards. In this case only anxiety and examiner could be studied as independent variables, using a two-by-two factorial design. The test data comprising dependent variables consisted of the following: (*a*) mean reaction time to all four cards; (*b*) difference between mean reaction time to the colored cards and that to the achromatic ones; (*c*) total number of responses; (*d*) total $F + \%$; (*e*) number of cards on which the first response had a poor form level; (*f*) difference between $F + \%$ for the colored cards and $F + \%$ for the achromatic ones; (*g*) number of whole responses; (*h*) movement; (*i*) shading; (*j*) *Sum C* scores; (*k*) Elizur's anxiety score; (*l*) animal content ($A + Ad$); (*m*) human content ($H + Hd$); (*n*) rejected cards; and (*o*) spontaneous, immediate remarks concerning the color. The first four of these scores were handled by analysis of variance while the remainder were analyzed using chi square and related procedures.

Reaction time. Much has been said in Rorschach literature concerning the psychological meaning of reaction time. A significant increase in reaction time to colored cards is generally considered an indication of "color shock" [11, 14]. Emotionally disturbed individuals will presumably have a long reaction time or show other disturbances due to colors, especially on Cards II and III. From this we might expect that, in general, mean reaction times for the colored cards would be longer than those for the achromatic versions. Moreover, there should be a significant interaction between color and anxiety with respect to reaction time if the above theory is correct.

A more comprehensive analysis was made of reaction time than of any other variable, mainly because of the continuous nature of reaction time scores, which permits the use of analysis of variance. Since the original scores were skewed in the typical fashion of reaction time data, an inverse square root transformation was employed to normalize the score distributions prior to analysis. An analysis of variance of the transformed reaction time scores for Trial 1 only and for Rorschach Cards II and III is presented in Table 2.

In this analysis, all insignificant, higher order interactions were pooled with the within sum of squares to provide an error term with 57

Table 2

Analysis of Variance of Transformed Reaction Time Scores for Rorschach Cards II and III, First Trial Only

Source of variation	Mean square	*df*	*F*	*p*
Examiner	220.9	1	1.63	—
Anxiety	11.4	1	.08	—
Inkblot	.8	1	.01	—
Color	876.2	1	6.50	.02
Examiner by color	581.1	1	4.31	.05
Examiner by anxiety by inkblot	1095.5	1	8.13	.01
Pooled residual	134.8	57		

df. Only one of the main effects was significant. Reaction time to the colored cards was significantly faster than response to the achromatic cards. Although this was more the case for Examiner Y than for Examiner X, as evidenced by the significant examiner-color interaction, these results suggest, contrary to the opinion of many clinicians, that bright red color, as applied on Cards II and III, evokes an earlier response than if these cards were achromatic.

The critical source of variance from the point of view of Rorschach theory, the interaction between anxiety and color, was negligible. Both "anxious" and "nonanxious" women responded in an identical manner. The hypothesis that differences in reaction time to colored and achromatic inkblots are related to anxiety has not been substantiated by analysis of response to the first card only.

The interpretation of the significant triple interaction, examiner-anxiety-card, is unclear. Assuming that this is a nonchance result, it probably reflects examiner differences more than anything else and would prove theoretically important only if one or more of the lower order effects involving these three variables had also been significant.

The same analysis of Cards VIII and IX revealed no significant factors related to reaction time. Unlike Cards II and III, reaction time to these cards is the same in both the multicolored, standard form and in the achromatic experimental version. These results are consistent with the findings of Dubrovner *et al.* [5], who found no differences in reactivity that could be attributed to color. The present study suggests, however, that the bright red color may stimulate the subject to respond more quickly.

On the other hand, Siipola [16] reported just the opposite for a number of colored areas of the Rorschach inkblots. Using matched colored and achromatic versions of smaller inkblots obtained by cutting out the usual detail areas of Cards II, III, VIII, IX, and X, she found consistently longer mean reaction times for college women when color was present than when it was absent. Additional experimentation, with color and inkblot varied systematically in an extensive factorial design, is necessary to resolve these apparent contradictions.

Table 3 contains the results for reaction time on Cards II and III when the first two cards presented to the subject are considered simultaneously. Only one source of variation, the order of presentation of color, proved significant. The colored-before-achromatic order yielded a significantly faster reaction time than the reverse order of presentation. Again it appears that the bright red colors are stimulating to the subject, this time shortening the latency of his response to the card which follows.

At the third level of analysis, employing all four cards, two scores were derived from the reaction time data: (*a*) mean reaction time; and (*b*) difference between mean reaction time to the colored cards and mean reaction time to the achromatic cards. The latter was of particular interest since previous writers [2, 11] on color shock have stated that a delay in reac-

Table 3

Analysis of Variance of Transformed Reaction Time Scores for Rorschach Cards II and III, First Two Trials

Source of variation	Mean square	*df*	*F*	*p*
Independent observations				
Anxiety	0	1	0.00	
Order of color	2080	1	7.07	.01
Order by anxiety	488	1	1.66	
Residual between subjects	294	60		
Correlated observations				
Trial	7	1	0.10	
Color	66	1	0.96	
Trial by anxiety	6	1	0.09	
Color by anxiety	149	1	2.16	
Residual within subjects	69	60		

tion time to a colored card is one sign of "shock" that presumably occurs with greater frequency in neurotic or anxious individuals than in others. An analysis of variance was made for each variable, breaking down the total variance into components attributable to examiner, anxiety, and examiner-anxiety interaction, each having 1 *df*, and an error term with 127 *df*. In neither case did either of the main effects or their interaction approach significance.

The results of these three different sets of findings concerning the extent to which reaction time is a function of examiner, anxiety, inkblot, color, and their interactions can be summarized as follows. Contrary to the opinion of many psychologists, there is little evidence to support the use of a long reaction time to a chromatic inkblot as a sign of color shock. Likewise, there is little support for the contention that anxious individuals take longer to respond to colored than to black and white inkblots, whereas nonanxious subjects do not. If anything, the bright red color on Cards II and III apparently stimulates most individuals to respond more quickly. This stimulation may carry over to the next card, as illustrated by a quicker response to the second card when the first inkblot was partly covered by bright red color. Throughout the analysis of reaction time a consistent difference was found between Cards II and III on the one hand and Cards VIII and IX on the other. This finding raises some interesting questions for further research concerning the comparative effects upon reaction time of variation in hue, configuration, and mode of applying the color to the inkblot.

Number of responses. Although the number of responses given to the first card ranged from one to fifteen, most women gave only one or two responses per card. Because of the discreteness and skewness of the data when only the first card was considered, it was necessary to resort to less efficient methods than analysis of variance to determine the role played by each independent variable. The distribution of scores was dichotomized close to the median, and fourfold tables were constructed; one independent variable was considered at a time and the others ignored. This was done separately for Cards II and III and for Cards VIII and IX as well as for all four inkblots combined. The only variable which showed a significant relationship to number of responses on the first card was color for Cards II and III. Inspection of Table 4 reveals that significantly more women gave more than one response to the colored cards than to the achromatic ones. Apparently the bright red colors on Cards II and III stimulate the individual to greater productivity and also to respond more quickly than otherwise.

Table 4

Number of Responses to the First Card as a Function of Color, Rorschach Cards II and III

Number of responses	Color of inkblot		Total
	Chromatic	Achromatic	
2 or more	22	13	35
1 or less	10	19	29
Total	32	32	64

$\chi^2 = 4.04$ ($p = .05$).

When the total number of responses for all four cards was considered, the frequency distribution was only moderately skewed. An analysis of variance using a two-by-two factorial design revealed no differences among examiners or high- and low-anxiety groups.

These negative results are contrary to those obtained by Eichler [7] and by Westrope [18]. Eichler reported a significant decrease in the number of responses given by individuals under stress conditions. However, in a study relating Rorschach variables to Taylor's anxiety scale, Westrope found just the opposite. Significantly more responses were given by her high-anxiety group than by the low-anxiety group.

The interaction between examiner and anxiety was also negligible. Since the distribution of scores was moderately skewed, an additional test was made by dichotomizing the distribution at the median (nine responses) and constructing fourfold tables for each independent variable. Again, no differences between groups were revealed. The results obtained in this experiment suggest that there is little relationship between the number of responses and manifest anxiety as measured by the MMPI. Although previous experimenters [4, 5, 13, 15] have concluded that color does not increase productivity,

the present findings indicate that the bright red hue on Cards II and III may result in a greater number of responses.

Form level. Each response was scored either "plus" or "minus," using Beck's tables [1] as a guide. When only responses to the first card were considered, it was discovered that no women gave minus responses to Card III, only 2 gave one or more minus responses to Card VIII, 4 to Card II, and 16 to Card IX. Since a balanced design had been used by which each card would be presented first to 32 different women, it is quite clear that Card IX is more likely to yield a minus response than any of the other inkblots used in this study. This is compatible with the often reported fact that Card IX has the longest reaction time and is the most frequently rejected card.

Using presence of one or more minus responses to the first card vs. no minus responses as one variable and one of the independent variables (examiner, anxiety, inkblot, color) as the other, fourfold tables were constructed and tested for significance. With the exception of the differences already noted from one inkblot to the next, none of the relationships even approached statistical significance. Similar results were obtained when $F + \%$ for the first card was analyzed.

Consideration of total $F + \%$ based upon all four cards yielded different results. A summary of the analysis of variance for the two-by-two factorial design is presented in Table 5. There was no appreciable difference in the mean $F + \%$ for the high- and low-anxiety groups. This is consistent with the negative results reported by Westrope (18), who also used criterion groups based on the Taylor anxiety scale, for $F + \%$ on the colored cards alone as well as for all ten cards. However, the subjects tested by Examiner Y gave a significantly greater percentage of responses with poor form level than those tested by Examiner X. Such results, though indicating the importance of examiner differences, are of little theoretical consequence since the examiner-anxiety interaction is negligible.

Two other variables dealing with form level were studied, the number of cards on which the first response had poor form level and the difference between $F + \%$ on the colored and achromatic cards. The latter was of particular interest because of the investigation by Williams [19] suggesting that drop in form level on colored cards is a sign of poor control and may also be related to anxiety. Neither variable proved related to either examiner or anxiety in the present study. Fifty-five per cent of the women in both the high- and low-anxiety groups gave all four first responses with good form level. Both groups also showed the same form level on the colored cards as they did on the achromatic versions. Contrary to the findings of Lazarus [12], there is no evidence to support the hypothesis that a drop in form level generally occurs in moving from a black and white inkblot to a colored one.

Table 5

Analysis of Variance for Total $F + \%$

Source of variation	Mean square	*df*	*F*	*p*
Examiner	1378	1	5.58	.02
Anxiety	43	1	0.18	—
Examiner by anxiety	8	1	0.03	—
Residual	247	124		

Other Rorschach variables. None of the remaining variables was sufficiently continuous in nature to permit the use of analysis of variance. Consequently, it was necessary to resort to categorical treatment, using for the most part fourfold tables to investigate the significance of the major factors under consideration.

The number of responses using the whole inkblot, the number employing movement as a determinant, and the number using shading were analyzed for the first card only as well as for the summary data based on all four cards. In addition, *Sum C,* Elizur's anxiety score, animal and human content, rejected cards, and occurrence of spontaneous, immediate remarks to the color were carefully investigated for any differences between examiners or between the high- and low-anxiety groups. Aside from minor differences between examiners in the number of movement responses, no significant factors were discovered for any of the remaining variables.

The negative results for Elizur's anxiety score are of special interest because of its reported validity as a measure of anxiety [8, 18]. The mean score for the high-anxiety group

was 1.26, while that for the low-anxiety group was 1.15, a negligible difference.

One explanation for the lack of significant results for Elizur's anxiety score, as well as for some of the other Rorschach variables, is that the use of only four cards, instead of the standard set of ten cards, results in scores of lower reliability. Reducing the number of cards administered to each subject may have lowered the reliability of Rorschach scores in the present study. On the other hand, the number of subjects employed is sufficiently large that a rather small relationship should prove statistically reliable. Although lack of such a relationship in a single experiment does not prove its nonexistence, it does raise serious questions concerning the uncritical adoption of the Rorschach score in question for use in the clinic.

Summary and Conclusions

This experiment was designed to investigate the relationships of anxiety, color, stimulus configuration, order of presentation, and examiner with a number of different Rorschach variables. High- and low-anxiety groups containing 64 subjects each were selected from 696 college women on the basis of MMPI scores for Taylor's Manifest Anxiety Scale and Winne's Scale of Neuroticism. Achromatic copies of Rorschach Cards II, III, VIII, and IX were carefully prepared and used together with corresponding standard cards. Four inkblots, two colored and two achromatic in a predetermined order and balanced design, were administered in a standard fashion to each subject by one of two examiners. The more experienced of the two examiners judged each subject as "anxious" or "nonanxious" at better than the .001 level of significance, solely on the basis of behavior during the individual testing session; this provided a check on the adequacy of the criterion.

The results and conclusions were as follows:

1. The colored versions of Cards II and III yielded significantly shorter reaction times and significantly more responses than did the achromatic versions of these same inkblots. However, no differences ascribable to color were obtained for these same variables on Cards VIII and IX. This finding suggests that the bright red color of Cards II and III is more stimulating than the multicolored pastel shades of Cards VIII and IX. The shorter reaction times to the colored versions of Cards II and III are contrary to expectations from color shock theory.

2. The colored-before-achromatic order of presentation resulted in a significantly shorter mean reaction time for Cards II and III than did the reverse order, achromatic before colored. This is consistent with the hypothesis that the bright red colors of Cards II and III stimulate the individual to respond more quickly.

3. Women tested by Examiner Y gave a significantly greater percentage of responses with poor form level than did women tested by Examiner X, indicating the importance of controlling the examiner as a variable in research of this type.

4. Contrary to the expectations of earlier Rorschach theory, none of the remaining Rorschach variables proved significantly related to anxiety, color, or order of presentation.

It is clear from the present research that many conceptions of how color, examiner, personality, and order-of-presentation variables affect performance on the Rorschach test will have to be revised considerably. Further studies similar to this one, which compare several factors simultaneously by use of modern experimental design, are needed to resolve apparent contradictions in the literature.

Received January 8, 1954.

References

1. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* (2nd Ed.) New York: Grune & Stratton, 1949.
2. Benton, A. L. The experimental validation of the Rorschach test II. The significance of Rorschach color responses. *Amer. J. Orthopsychiat.*, 1952, **22**, 755-763.
3. Brosin, H. W., & Fromm, E. G. Rorschach and color blindness. *Rorschach Res. Exch.*, 1949, **4**, 39-70.
4. Buker, S. L., & Williams, M. Color as a determinant of responsiveness to Rorschach cards in schizophrenia. *J. consult. Psychol.*, 1951, **15**, 196-202.
5. Dubrovner, R., Von Lackum, W., & Jost, H. A study of the effect of color on productivity and reaction time in the Rorschach test. *J. clin. Psychol.*, 1950, **6**, 331-336.

6. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
7. Eichler, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.*, 1951, **46**, 344-355.
8. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach Res. Exch.*, 1949, **13**, 247-284.
9. Hertz, Marguerite R. The Rorschach: thirty years after. In D. Brower & L. Abt (Eds.), *Progress in clinical psychology.* Vol. I, Section I. New York: Grune & Stratton, 1953. Pp. 108-148.
10. Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853-855.
11. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
12. Lazarus, R. S. The influence of color on the protocol of the Rorschach test. *J. abnorm. soc. Psychol.*, 1949, **44**, 505-516.
13. Perlman, J. A. Color and the validity of the Rorschach 8-9-10 per cent. *J. consult. Psychol.*, 1951, **15**, 122-126.
14. Rapaport, D. *Diagnostic psychological testing.* Vol. II. Chicago: Year Book Publishers, 1945.
15. Sappenfield, B. R., & Buker, S. L. Validity of the Rorschach 8-9-10 per cent as an indicator of responsiveness to color. *J. consult. Psychol.*, 1949, **13**, 268-271.
16. Siipola, Elsa. The influence of color on reactions to ink blots. *J. Pers.*, 1950, **18**, 358-382.
17. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
18. Westrope, Martha R. Relations among Rorschach indices, manifest anxiety, and performance under stress. *J. abnorm. soc. Psychol.*, 1953, **48**, 515-524.
19. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21-29.
20. Winne, J. F. A scale of neuroticism. An adaptation of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1951, **7**, 117-122.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Motor and Fantasy Correlates of Rorschach Human Movement Responses

Jerome L. Singer
Franklin D. Roosevelt VA Hospital, Montrose, N.Y.[1]
and Jack Herman
New York University

One of the most unusual and theoretically interesting features of Rorschach's inkblot test is the human movement determinant (*M*). In linking the individual's perception of humans in action on the blots to stable motility, on the one hand, and imaginativeness or inner living, on the other, Rorschach [11] was in essence providing a concrete formulation of a theory of vicarious functioning similar to that presented by a number of writers in diverse fields of psychology and psychoanalysis [4, 6, 9, 18]. A series of investigations has provided data supporting Rorschach's linkage of motor inhibition and *M* responses [7, 14, 15, 19]. There is also evidence that *M* responses are associated with measures of fantasy or imaginative tendencies [1, 2, 8, 12, 13, 16]. Two earlier studies in this series have reported positive correlations between the number of *M*[2] and the ability of *S*s to delay their speed of writing. A negative correlation between *M* and the degree of spontaneous motor activity manifested during a waiting period also emerged [7, 15].

The objectives of the present investigation included a partial replication of earlier findings [15] with another sample and an extension of the behavioral situations to include measures of fantasy as well as motor activity. The specific hypotheses are:

1. Schizophrenic *S*s assigned to a High-*M* group on the basis of their Rorschach records differ significantly from Low-*M* *S*s on a series of tasks measuring motor activity and fantasy behavior or introspectiveness. High-*M* *S*s show greater delaying capacity, less impulsivity, less spontaneous motor response, and higher scores on measures of fantasy. A subhypothesis in this connection was derived from an earlier study [15] in which it was found that waiting-room activity scores varied not only with *M* [illegible] the relationship of *M* and the weighted responses (*Sum C*). This finding offering support to Rorschach's juxtaposition these two types of response determinants (*Sum C*) as an index of relative weight of extratensive or extratensive response tendencies. It was hypothesized, therefore, that, at least in the case of spontaneous motor activity (waiting-room behavior), the degree of activity varies with the increasing weight of color in the experience balance of the patients.

2. If Rorschach descriptions of the interrelations of motor and fantasy activity are correct, one may anticipate a certain consistency in *S*s' performance from one situation to the next. Individuals with relatively numerous *M* also manifest greater motor inhibition, less spontaneous motor activity, and more fantasy. Specifically, the subtests that tap various aspects of the motor inhibition – motion perception–fantasy dimension are intercorrelated and define a common factor or meet some appropriate standard of interrelatedness. These correlations may not be high because of the numer-

[1] Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions do not necessarily reflect the policy or opinion of the Veterans Administration. The authors are indebted to Drs. Harold Wilensky and Jacob Cohen for helpful criticism, and to staff and psychological trainees at this station for assistance in various phases of the study.

[2] The focus in this study was on human movement alone. Previous studies have yielded equivocal results concerning the relationship of animal movement (*FM*) and inanimate movement (*m*) to *M*. This empirical distinction between the types of movement responses merits considerable research exploration.

ous situational, uncontrolled factors that undoubtedly limit the reliability of a Rorschach examination. Some gross consistency of performance is certainly to be expected, however, if Rorschach's formulation has any validity. When Rorschach factors are used in relative isolation in this investigation with little concern for "configurational" features, any significant correlations between *M* and behavioral situations should be all the more striking as evidence of the essential meaning of the human movement determinant.

Subjects

In order to repeat some of the conditions of the previous study [15], a new sample of 60 male schizophrenic veterans was employed. The *S*s received clinically administered Rorschachs which were rescored in consultation by experienced examiners. After administration of the five experimental procedures, the total sample was divided into High- and Low-*M* subgroups. The criteria derived from examination of Rorschach score distributions for the sample were more than one *M* (high *M*) and a *Sum C* of 2.5 or more (high *Sum C*). The High- and Low-*M* groups did not differ significantly in IQ (Wechsler-Bellevue Scale), age, educational level, or other Rorschach variables except *M*. Although differences in total responses (*R*) between the two main groups were larger than in the previous study, most of this difference could be attributed to differences in *M* and *FM*, the mean *M* of the High-*M* *S*s being roughly eight times as great as that of the Low-*M* group. Somewhat in keeping with a finding reported by Wittenborn [20], the pattern of color responses was somewhat different for the two groups, although not significantly so. Thus *FC* responses tended to correlate somewhat better with *M* and *R* than with the other color determinants.

On the whole the characteristics of this sample did not differ markedly from those of the group of 50 schizophrenics described in the earlier publication. Diagnostic labels, degree of morbidity, or duration of illness did not seem to vary systematically. Again as noted in the earlier investigation, the "dilated" experience-type group (high *M*: high *Sum C*) proved to be somewhat more generally productive and "interesting," albeit no less disturbed than the other subgroups.

Procedure

Each *S* participated in five experimental situations in more or less the same order. Occasional changes in sequence were necessitated when patients became disturbed or showed wandering attention. In all cases the examiner (always one of the authors) was unfamiliar with patients' Rorschach scores to minimize any unconscious influence on performance.

1. *Waiting-room activity ratings.* This procedure, modified from a method described by Rickers-Ovsiankina [10], was described in a previous paper [15]. The *S* was conducted into a waiting room and informed that *E* would call for him shortly. He was seated near a table on which were placed a variety of objects and games. The *S* was then observed for 15 minutes by means of a one-way vision screen. All activity and movements were recorded at 30-second intervals. These protocols were subsequently rated by judges on a six-point scale for degree of activity and environmental responsiveness. The scale itself was simple to use and had reasonably high interrater reliability [15]. A patient who sat quite still except for an occasional shifting of weight or puff on a cigarette would obtain in minimal score, while an active individual who paced the floor, manipulated many of the objects on the table, and spoke aloud would obtain a maximum score.

2. *Motor inhibition task.* This was a simple test modified from Downey's Will-Temperament Scale and employed in previous investigations of this series [7, 15]. The *S* was required to write the phrase "New Jersey Chamber of Commerce" as slowly as he possibly could without stopping motion or lifting the pencil from the page. A sample of writing at normal tempo was obtained prior to the experimental trial to evaluate the extent of *S*'s effort to comply with the instructions. This precaution was necessary because of the fairly disturbed patient group employed. The data here were the times in seconds taken to write the phrase.

3. *Thematic Apperception Test.* Cards 1, 2, 4, 6BM, and 7BM of the TAT were administered under two conditions. Initially each *S* was asked to describe each card. After com-

pletion of the five descriptions, the cards were presented again with the usual instructions to make up a story. The protocols were scored by means of a modification of Weisskopf's transcendence index [17], a measure of the extent to which content is introduced into *S*'s story which goes beyond the actual material inherent in the stimulus. Additional characters, emotions attributed to characters, mention of past or future events were all scored as transcendent items according to a system of thought units. A total score for each series was obtained. In order to derive a measure of voluntary fantasy the difference between the transcendence scores for the *description* and *story* conditions was employed. While other investigators have employed more global ratings of over-all creativity of themes [13, 16], it was felt here that Weisskopf's transcendence index would be most easily quantified and reliably rated. This test situation was felt to be a good initial approximation of an individual's capacity to engage in imaginative behavior, manipulating situations, persons, and events "in his mind's eye," so to speak. The *story* instructions in effect demand of *S* that he adopt an attitude toward the "possible," as Goldstein might put it, by introducing situations and persons not directly called forth by the card itself. Actual verbal fluency is not sufficient here, although it bears some relation to the task. Thus, a person might present a lengthy, obsessive description of the card and of the specific clothes or position of the characters yet never obtain a single point for transcendence. A story like: "This boy's dreaming of being a great violinist. If his parents help him and his teacher encourages him he'll practice hard day by day and finally on an evening 20 years from now he'll win fame before a Carnegie Hall audience," on the other hand, while relatively short, has a relatively high transcendence score.

4. *Cancellation test.* This is a very simple procedure based on an unpublished test devised by Dr. Seymour Klebanoff. Rows of letters in randomized sequence on a printed sheet are presented to *S* who is required to encircle all *o*'s and *e*'s. He is told to go as rapidly as he can but be as accurate as possible. The extreme simplicity of the task makes it possible to obtain a measure of impulsivity by noting errors and time, since the task makes little intellectual demand on *S*s. This test was included chiefly in an exploratory attempt to observe whether *S*s with low *M* and high *Sum C* would show many errors. Perhaps more complex situations might have been more discriminating, but preliminary research with similar patients revealed that such tasks made considerable demands on the limited attention of schizophrenics.

5. *Guilford's Thinking Introversion Scale (Factor T of the* Inventory of Factors S T D C R). The items on this scale are described by Guilford as defining a factor which involves "an inclination to meditative or reflective thinking, philosophizing, analysis of one's self and others, versus an extrovertive orientation" [5, p. 000]. Certainly this description resembles some of the propensities of High-*M* individuals as described in the Rorschach literature. While the inventory was developed with college students and has some of the usual limitations of questionnaires, it was included here in an exploratory way to obtain some measure of the patient's self-perception as a thinking introvert. In the interests of maintaining a reasonable time limit and feasible battery of tests for this schizophrenic group, only the T scale of the inventory was employed. To facilitate administration to this group, items were printed on cards and sorted in the manner of the MMPI. It must be recognized that these deviations from the standardized mode of administration may well alter the test and increase uncontrolled variability.

Results

Table 1 presents pertinent background information, Rorschach scores, and results on the experimental situations for the High- and Low-*M* groups. Inspection of this table reveals that the general trend of the data is in the hypothesized direction. High-*M* *S*s differ from the Low-*M* patients by manifesting longer inhibition times in the writing task, less spontaneous activity in the waiting room, greater transcendence of stimulus content in the five TAT cards, fewer errors in the cancellation task,[8] and greater tendency toward thinking

[8] The groups did not differ significantly in speed of performance although Low-*M* *S*s tended to work a little more rapidly.

introversion on the Guilford scale. Analysis of variance for the motor inhibition times reveals an F ratio of 9.6, significant beyond the .01 level, with the group differences also significant by the t test beyond .01. Waiting-room scores also yielded an F ratio of 5.19, significant beyond the .01 level, with the group means differing significantly by the t test at the .01 level. Scores on the transcendence index yielded an F ratio of 5.96, significant at the .01 level, and again a t test yielded significance beyond .01 for the difference in group means. Neither the cancellation errors nor the Guilford scores yielded significant t or F values although trends were evident. Results of these tests, therefore, tend on the whole to support the initial hypothesis.

Table 1

Means for High- and Low-*M* Groups on Pertinent Background Data, Rorschach Variables, and Experimental Situations

Matching variables	Groups	
	High-*M* group ($N = 28$)	Low-*M* group ($N = 32$)
Background variables		
Age	29.6	31.4
Education (years)	11.7	11.2
IQ (Wechsler-Bellevue)	111.5	108.2
Rorschach Variables		
M	2.6	0.3**
Sum C	2.8	2.7
FM	4.7	2.1
Shading	2.6	2.3
Total responses	25.3	17.3
Experimental situations		
Motor inhibition time	172.21	62.97**
Waiting-room activity score	1.92	3.13**
TAT transcendence index	21.96	12.69**
Cancellation test errors	3.00	4.25
Guilford Factor T scale	39.42	36.25

**Differences between groups significant beyond the .01 level.

As to the specific prediction concerning the spontaneous motor activity in the waiting room, Table 2 reveals a regular increase in activity scores as the groups change from high *M*: low *Sum C* through low *M*: high *Sum C*. This result is strikingly similar in trend and in actual scores to the results obtained with the population in the previous study [15]. This relationship between overt response and the *M*: *Sum C* ratio emerged only for the waiting-room scores.

Table 2

Means of Waiting-Room Activity Ratings for High- and Low-*M* *S*s Subdivided by *Sum C* Scores

Group	*N*	Rorschach variables		Waiting-room activity ratings
		M	*Sum C*	
High *M*: Low *C*	14	2.3	1.2	1.58
High *M*: High *C*	14	3.2	4.9	2.29
Low *M*: Low *C*	18	0.4	1.2	2.89
Low *M*: High *C*	14	0.3	5.2	3.43

The next phase of the analysis of results involved testing the hypothesis concerning the consistency of *S*s from situation to situation. Two methods for ascertaining whether a common factor underlies *M* and the five experimental situations were employed. The first of these was Blomqvist's test of relatedness [3], a nonparametric method which provides a t test of the significance of concomitant variation of two or more tests for a common population. It involves dichotomization of all test scores into high and low and makes no assumptions concerning the distributions. For this reason it is useful with Rorschach data and particularly where the range of scores is quite limited, as in the case of the *M* response in this group of schizophrenics. The technique does not, of course, provide the precise information available from a factor analysis which affords an estimate of the degree of relationship and of the factor loadings of the various subtests.

Applying Blomqvist's test to the distribution of *M*, treated as one of a battery of tests, and to the five experimental situations results in a t of 8.48, significant beyond the .001 level. This finding suggests that the six sets of scores are indeed grossly related. To evaluate the extent to which all experimental situations share in this concomitant variation, Blomqvist's test was applied to the six sets of scores in pairs yielding the matrix presented in association with p values in Table 3. It is apparent from inspection of this table that a common factor must indeed subsume *M* and these seemingly diverse test situations. Only the Guilford scale

fails to show uniformly significant p values for relatedness with the other tests.

Table 3

Matrix of p Values for Paired Experimental Situations Based upon Application of Blomqvist's Test of Relatedness

	1	2	3	4	5	6
1. *M*		<.01	<.01	<.05	<.01	>.01
2. Motor inhibition	<.01		<.01	<.05	<.05	>.20
3. Waiting room	<.01	<.01		<.01	<.05	>.20
4. Transcendence	<.05	<.05	<.01		<.01	>.20
5. Cancellation errors	<.01	<.05	<.05	<.01		<.01
6. Guilford's Factor T	<.01	>.20	>.20	>.20	<.01	

The application of Blomqvist's test would thus appear to offer support to the second general hypothesis of this investigation. The next step in evaluation of the data involved the calculation of product-moment correlations between Rorschach variables and the experimental situations. This procedure was employed with some trepidation because of the unknown quality of Rorschach distributions, the extremely restricted range of *M* and other variables in this sample, the limited *N*, and, finally, the inevitable interrelationship of *M* and the total responses (*R*), which cannot legitimately be eliminated by partial correlation since the two variables are not experimentally independent. In line with Wittenborn's reasoning [20], however, the extraction of a centroid factor by Thurstone's multiple factor analysis was carried out in a general attempt to test the hypothesis of commonality. The advantage of this method is that some indication of extent of association could be obtained. Other Rorschach variables bearing on the common factor might also be expected to emerge.

Calculation of Pearson product-moment *r*'s for IQ, variables on the Rorschach, and the five experimental situations revealed relatively few significant correlations with the exception of the matrix of *M* and three of the experimental situations, motor inhibition, waiting room, and the TAT transcendence index. Other variables like *FM*, the Guilford, and cancellation error scores showed trends in the expected direction, but the correlation coefficients did not reach significance. Extraction of centroid factors for those variables which, following a cluster analysis, showed sufficient significant interrelations revealed one clear factor which had its highest loadings from *M* and the motor inhibition times (Table 4). These findings tend to offer partial support to the second hypothesis. It is hoped that further research with larger groups and normal *S*s who may be expected to show a greater range of *M* responses will provide a more suitable test of this hypothesis using the factor analytic approach.

Table 4

Matrix of Intercorrelations for *M* and Three Experimental Variables

Variable	1	2	3	4
1. *M*		.539	−.270	.347
2. Motor inhibition	.539		−.244	.293
3. Waiting-room activity	−.270	−.244		−.077
4. TAT transcendence index	.347	.293	−.077	
Centroid factor loadings	.749	.714	.376	.465

A final use of correlation was attempted to evaluate the role of the color variables. Although *Sum C* and the separate color scores did not yield any significant relationships with other variables, inspection of Table 2 suggests that at least for waiting-room activity ratings some correlation between the *M* : *Sum C* ratio and experimental situations must exist. Scores from 1 to 4 were assigned to the High-*M* : Low-*Sum C*, High-*M* : High-*Sum C*, Low-*M* : Low-*Sum C*, and Low-*M* : High-*Sum C* groups in that order. Calculation of product-moment *r*'s for this *M* : *Sum C* score and the experimental situations revealed no changes or lowering of their correlations with *M* alone. The correlation for the waiting-room scores and this variable was + .47, almost twice that obtained when *M* alone was involved. Thus there seems to be here a hint of validity to Rorschach's linkage of movement and color, and further exploration with other behavioral correlates clearly seems warranted.

Discussion

The data of this investigation on the whole

lend support to the view that a person's capacity or tendency to control overt motor activity is related to his capacity for imaginative or fantasy activity. Rorschach's human movement determinant may indeed be a gross but valid measure of these alternative response tendencies. While specific predictions in the individual case are not warranted from the results presented here, it is clear that this motor inhibition—fantasy dimension merits more extensive exploration. Should further experimentation extend the scope of this triadic relationship of motion perception, motor response, and imagination to other behavioral situations or to other populations, a new approach to understanding an important phase of personality development may be opened. Can it be shown that control of motor response, or, more generally, postponement of gratification is accompanied by increased activity on what Lewin [6] has termed the "level of irreality"? At least in moderation such control has great functional value for man. The ability to defer immediate motor response directed toward need gratification with the consequent resort to the realm of fantasy gives man a control of his future through imagery and planning. Indeed such a pattern of postponement of immediate effort and evolution of a timeless behavioral space available for what Freud [4] has called "experimental action" may well be one of the key phases of maturation and, more broadly, civilization. It may also make possible a medium for experiencing anxiety unknown to animals or primitive man. Rorschach's *M* response is significant insofar as it serves as a measure of presumably long-standing personality characteristic reflecting this pattern of motor control and enhanced fantasy. Perhaps in a more refined and readily quantifiable form[4] this threshold for perception of humans in motion may serve as a valuable clue to important response tendencies that may differentiate not only pathological groups but diverse subcultures within our society. The "deferred gratification pattern" often represented as characteristic of American middle-class values might well be explored by tests of the type used here or more complex extensions of them. Of interest might be an attempt to link early parental training or perceptions of familial figures to the development of the *M* response and its behavioral correlates.

[4] A promising attempt in this connection was reported recently by Barron [2].

Summary and Conclusions

Previous studies in this series had demonstrated significant relationships between Rorschach human movement responses and measures of motor inhibition and spontaneous motor activity, thus affording partial support for Rorschach's postulated relationship of overt motor activity, perception of motion, and fantasy. The present investigation was designed as a partial replication of earlier work and an extension to include measures of fantasy or imaginative capacity as well as motor activity. It was hypothesized that *S*s differing in Rorschach *M* productivity would differ on tasks tapping motor control and fantasy. It was also anticipated that they would manifest rough consistency in their performance on the motor and fantasy situations in support of the general theory of vicarious motor, sensory, and fantasy functioning.

Sixty male schizophrenic patients who had recently been administered Rorschachs underwent five experimental procedures. The *S*s were divided into High- and Low-*M* groups on the basis of Rorschach protocols, two *M* or more being the dividing point. High-*M* *S*s ($N = 28$) did not differ significantly from Low-*M* *S*s ($N = 32$) in age, diagnosis, education, IQ, or other pertinent Rorschach variables. The experimental situations were: (*a*) observation of *S*s alone in a waiting room with ratings of the degree of activity manifested; (*b*) a motor inhibition task, ability to write a phrase as slowly as possible; (*c*) Thematic Apperception Test stories scored for transcendence, i.e., introduction of content not actually represented in the stimulus; (*d*) Klebanoff cancellation test scored for errors; (*e*) Guilford's scale for the thinking introversion factor (T) from the *Inventory of Factors STDCR*. Analyses of variance and *t* tests were employed to evaluate significance for the individual tests. Correlational procedures and Blomqvist's nonparametric test of relatedness were employed to evaluate consistency of performance from test to test.

Results on the whole supported both general

hypotheses. High-*M* *S*s differed significantly from Low-*M* *S*s by manifesting longer motor delaying capacity, less spontaneous activity during the waiting period, and more frequent introduction of characters or incidents not directly represented in TAT stimuli. They showed nonsignificant trends to make fewer cancellation errors and to present greater self-description as thinking introverts than did Low-*M* patients. For the waiting-room spontaneous activity, the relationship of Rorschach color and movement proved to be particularly important, replicating a finding in an earlier study. With Blomqvist's test, a significant interrelationship of *M* and the five experimental situations emerged, suggesting the presence of a common factor underlying the six sets of scores. Extraction of a centroid factor from a matrix of those intercorrelations which reached significance revealed fairly high factor loadings for *M*, waiting-room activity scores, motor inhibition, and TAT transcendence. The factor analytic procedure, although based on not very tenable statistical assumptions for this sample and these distributions, also appears to lend at least partial support to the hypothesis that a common dynamic factor underlies motor control, motion perception, and fantasy behavior.

Received January 14, 1954.

References

1. Barrel, R. Subcategories of Rorschach human movement responses: a classification system and some experimental results. *J. consult. Psychol.*, 1953, **17**, 254-260.
2. Barron, F. Threshold for perception of human movement in inkblots and its personality correlates. *Amer. Psychologist*, 1953, **8**, 317-318. (Abstract)
3. Blomqvist, N. Some tests based on dichotomization. *Ann. math. Statist.*, 1951, **22**, 362-371.
4. Freud, S. *Collected papers.* Vol. IV. London: Hogarth, 1946.
5. Guilford, J. P. *An inventory of factors S T D C R.* Manual (Rev. Ed.). Beverly Hills, Calif.: Sheridan Supply Co., 1940.
6. Lewin, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.
7. Meltzoff, J., Singer, J. L., & Korchin, S. J. Motor inhibition and Rorschach movement responses: a test of the sensory-tonic theory. *J. Pers.*, 1953, **21**, 400-410.
8. Pickering, W. D. A comparison of predominant verbal levels on the Thematic Apperception Test with the Rorschach experience-balance. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1950.
9. Rapaport, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.
10. Rickers-Ovsiankina, Maria. Studies of the personality structure of schizophrenic individuals. I. *J. gen. Psychol.*, 1937, **16**, 153-178.
11. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
12. Schumer, Florence. Some behavioral correlates of Rorschach human movement responses. Unpublished doctor's dissertation, Yale Univer., 1949.
13. Shatin, L. Relationships between the Rorschach test and the Thematic Apperception Test. Unpublished doctor's dissertation, Harvard Univer., 1950-51.
14. Singer, J. L., Meltzoff, J., & Goldman, G. D. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.*, 1952, **16**, 359-364.
15. Singer, J. L., & Spohn, H. E. Some behavioral correlates of Rorschach's experience-type. *J. consult. Psychol.*, 1954, **18**, 1-9.
16. Vernier, Claire M., & Kendig, Isabelle V. Analysis of the relationships between various measures of creative productivity in two projective tests. Paper read at Amer. Psychol. Ass., Chicago, 1951.
17. Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379-390.
18. Werner, H. Motion and motion perception: a theory of vicarious functioning. *J. Psychol.*, 1945, **19**, 317-327.
19. Werner, H. Perceptual behavior of brain-injured mentally-deficient children. *Genet. Psychol. Monogr.*, 1945, **31**, 51-110.
20. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, **14**, 261-267.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

The Ammons FRPV Test and the WISC for Remedial Reading Cases[1]

Louis M. Smith and Arline R. Fillmore
Psycho-Educational Clinic, University of Minnesota

The Ammons Full Range Picture Vocabulary Test and the Wechsler Intelligence Scale for Children were given to 91 children diagnosed as reading disability cases at the Psycho-Educational Clinic. A reading disability was defined as a reading grade level, on a standardized achievement test, at least one year below the mental grade level on either the WISC or the Stanford-Binet. Eight of the subjects were girls and 82 were boys. The mean chronological age was 10.9 years, the mean reading grade was 3.2 years, and the mean intelligence level on the WISC was 101. Most of the children received both tests in the initial diagnostic interview, but some were given the Ammons while they were enrolled in the summer reading school.

Form A and Form B of the Ammons were comparable in mean and standard deviation. When combined the mean was 107.89, the standard deviation was 13.70. The mean of the WISC Verbal Scale was 99.66, Performance Scale 101.87, and Full Scale 100.85. The standard deviations were respectively 10.87, 12.51, and 10.71. The correlation between Form A and Form B of the Ammons was .90. The two forms were combined into a single scale and then correlated with the WISC scales. The correlations were .73 with the Verbal Scale, .54 with the Performance Scale, .75 with the Full Scale, and .61 with the scaled score of the Vocabulary subtest. When the correlation between the combined forms of the Ammons and the WISC Full Scale was corrected for the restriction of range on the WISC, the value was raised to .82.

In conclusion, the present writers feel that the Ammons is a useful screening test of intelligence for children with reading handicaps. The differences in mean scores indicate a possible limitation. However, the full significance of this cannot be fully weighed at the present time. Krugman, *et al.* [1] reported differences between the WISC and the Stanford-Binet similar in magnitude and direction to the present differences between the WISC and the Ammons.

[1]An extended report of this study may be obtained without charge from Louis M. Smith and Arline R. Fillmore, Psycho-Educational Clinic, University of Minnesota, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4321 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Brief Report
Received June 2, 1954.

Reference

1. Krugman, Judith I., Justman, J., Wrightstone, J. W., & Krugman, M. Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. *J. consult. Psychol.*, 1951, 15, 475-483.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

A *P* Scale for the Rorschach: A Methodological Study

Philip H. DuBois and Arnold H. Hilden
Washington University

The present study represents the application of a relatively new method of item analysis [3, 6] to Rorschach data. It can be readily shown that the total variance of a test can be divided into two parts: the sum of the item variances and the sum of all the covariances among the items. The total variance of a test with zero intercorrelations among the items, and hence zero interitem covariances, would be the sum of the item variances. Each item would then have nothing in common with the other items. The test would be completely heterogeneous, measuring different qualities rather than a single quality. It would be impossible to give a name to such a test, unless it predicted a criterion similarly heterogeneous. In that case its unity would come from without rather than from within.

In general, a psychometric test should be designed to measure a unitary characteristic, revealed by positive intercorrelations among the items. If these correlations are too high and approach unity, the test makes too few discriminations to be of value. With the type of material used in the present study, however, this difficulty was not encountered. Our procedure for selecting items was to maximize the saturation, defined as the ratio of the sum of the item covariances to the total variance of the test. This formula may be written as follows:

$$S = \frac{2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} C_{ij}}{\sigma^2_x} = \frac{\sigma^2_x - \Sigma\sigma^2_i}{\sigma^2_x} = 1 - \frac{\Sigma\sigma^2_i}{\sigma^2_x},$$

in which C_{ij} represents the covariance between any two items, and σ^2_x is the total variance.

It is readily apparent that maximizing the saturation is essentially the same operation as maximizing the internal consistency as estimated by Kuder-Richardson Formula 20. We can convert a saturation coefficient into r_{tt} by multiplying by $\frac{n}{n-1}$ in which n is the number of items.

It occurred to us that this procedure offered promise as a method of scaling Rorschach variables. Rorschach literature, while voluminous, has not drawn heavily on conventional psychometric methods.

The popular response, *P*, was defined by Rorschach [8, p. 359] as follows: "The vulgar answers represent the share in the collective or common method of sensing anything." Subsequently, a number of authors have expressed various definitions or rationales underlying the popular concept. Hertz [4], Beck [1, 2], Klopfer [5], and Rapaport [7], among others, have proposed different *P* lists. All, including Rorschach, define *P* on a definitely statistical basis. It is one of the few Rorschach categories so defined. There is, however, variation as to the exact statistical criterion. Klopfer says [5, p. 177] that "the term popular refers to . . . responses given most frequently . . . to a particular blot area" and "it seems most reasonable to use the total number of subjects . . . as a basis for determining the frequency with which a concept has been selected." He states that since a frequency tabulation of universal validity is not available, he has substituted a selected sample of popular responses which meet four qualifications: (*a*) they must occur frequently; (*b*) frequently used determinants must be present; (*c*) any "normal" person should be able to accept them; and (*d*) a *D* or *W* must be involved.

Beck insists that P is based on form alone. Rapaport gives no exact criterion, but states that "they must occur in at least every 4 or 5 records" [7, p. 313]. Hertz [4] uses a definite statistical basis. Beck [2] has used the most definite and elaborate basis, and has made the most exhaustive attempt to arrive at criteria for establishment of P responses.

In the present study the lists proposed by these four authors were taken as representative of populars used in clinical practice in this country. The four lists were combined into a composite list of 47 items, many of which were alternates or variations of others on the list. While the four authors are not in complete agreement on any item, the list was drawn up so that each of the four authors was fully represented.

Subjects

The subjects consisted of 200 normal adults, normal in the sense that they were not clinic patients. Most of them were tested by graduate students enrolled in a projective technique course, while the others were tested by one of the authors.

The population, consisting of 100 males and 100 females, was randomly divided into two groups, Group A, an experimental group, and Group B, a cross-validation group. Mean ages were 31.6 and 30.4 years, respectively.

Procedure

For Group A, the entire matrix of item variances and covariances for 47 items was computed. Saturation and internal consistency reliability coefficients are reported in Table 1 for the scales proposed by the four authors, in the two groups of 100 each and for the combined group of 200. For these scales, variation in saturation and reliability from Group A to Group B comes merely from sampling, since no maximizing is involved. Of the four scales proposed, Beck's has the highest internal consistency, .47, Klopfer's the lowest, .22. Even with Beck's scale, however, the internal consistency is lower than we would wish to have in a good psychometric device measuring a quantifiable aspect of human behavior along a continuum.

In order to develop a homogeneous scale of Rorschach populars, the entire collection of 47 items was considered as a source of possibilities. However, a number of the items, as indicated in Table 2, were considered as overlapping, representing definitions by different authors of the same general idea. Accordingly, only one of the items in such a group of "overlapping" was considered for inclusion at any one time.

Table 1

Saturations and Reliabilities (by Kuder-Richardson Formula 20) of the Several P Scales

		Group A		Group B		Combined	
Author	n	S	r_{tt}	S	r_{tt}	S	r_{tt}
Beck	21	.48	.51	.41	.43	.45	.47
Hertz	20	.36	.38	.39	.41	.38	.40
Klopfer	10	.15	.17	.24	.27	.20	.22
Rapaport	13	.24	.26	.35	.38	.30	.32
Saturation	18	.63	.67	.51	.54		

A nucleus of four items having the highest available interitem covariances was selected as a starting point. This nucleus consisted of items 1, 13, 19, and 23, as listed in Table 2. At each stage in the development of the saturation scale, all items were tested for inclusion in or exclusion from the nucleus.

It has been shown [6] that any item, j, not included in the nucleus, will add to the saturation if

$$\frac{2C_{zj} + \sigma^2_j}{\sigma^2_j} > \frac{\sigma^2_z}{\Sigma\sigma^2_i}.$$

On the other hand any item, k, included in the nucleus, should be removed if

$$\frac{\sigma^2_z}{\Sigma\sigma^2_i} > \frac{2C_{zk} - \sigma^2_k}{\sigma^2_k}.$$

A peculiar problem was encountered in the construction of the present scale in the matter of alternative possibilities, that is, different definitions of the same general idea. Accordingly, a formula was developed to determine whether any specific interchange would add to the saturation.

If j is any item in the nucleus and k is any item outside the nucleus, the saturation will be increased if the following inequality holds:

$$\sigma^2_z\,(\sigma^2_j - \sigma^2_k) > \Sigma\sigma^2_i\,(2C_{jk} + 2C_{zj} - 2C_{zk} - \sigma^2_j - \sigma^2_k).$$

Table 2

Composite *P* List

Item no.	Percentage passed** A	Percentage passed** B	Card and location†	Source‡	Content of the popular responses
1§	58	55	I, W	BHKR	Bat, butterfly, moth (R omits moth)
2	4	1	I, *W*	H	Bird (also item 1)
3	2	5	I, *W*	K	Winged creature (also item 1)
4	61	61	I, *W*	K	Items 1, 2, and 3 combined (generic concept)
5	1	0	I, *W*	H	Face of a person (not animal or cat)
6*	10	11	I, *D*3	B	Lower part of body (*D*4 if *D*3 attended to)
7	18	21	II, *W*, ½	BH R	Human being
8	0	1	II, *W*, ½	R	Bear, dog
9*	31	33	II, *D*1	B R	Bear, dog
10	32	40	II, *D*1, 7	H	Bear, dog, cow, animal
11	33	35	II, *D*1	K	Bear, dog, rabbit, bull, rhinoceros variety
12*	17	17	II, *D*3	B	Butterfly, moth
13*	68	83	III, *D*1,9	B R	Human beings (with or without *M*)
14	66	79	III, *D*1,9	HK	Human beings (only if *M* is present)
15	21	20	III, *D*3	B	Bow, bow-tie, hair ribbon
16¶	32	36	III, *D*3	HKR	Butterfly, or item 15
17	46	35	IV, *W*	BH R	Animal skin
18	7	17	IV, *D*2,6	H	Boot, foot, shoe
19*	77	86	V, *W*	BHKR	Bat, butterfly, moth (R omits moth)
20	4	6	V, *W*	H	Bird (also item 20)
21	90	92	V, *W*	K	Winged creature, or items 19 or 20
22*	12	17	V, *D*1	B	Leg, animal or human
23	51	65	VI, *W*, *D*1	BH R	Animal skin (*F* only or with shading)
24*	37	46	VI, *W*, *D*1	K	Animal skin, only when shading is present
25*	51	51	VII, *D*1	BH R	Face or head of human (also if *D*2, ½, *W* used)
26	69	67	VIII, *D*1	B	Animal (Beck's list)
27	89	83	VIII, *D*1	H	Animal (Hertz's list)
28	51	50	VIII, *D*1	K	Four-legged animal, with *FM*
29	93	90	VIII, *D*1	R	Animal
30*	21	24	VIII, *D*3	BH	Vertebrae, ribs, bone structure
31	6	9	VIII, *D*4	BH	Tree, bush
32*	19	20	IX, *D*3	B R	Human being
33	14	17	IX, *D*4	B	Face, head; of a man
34*	16	17	IX, *D*4	H	Face, head; person, not female
35*	16	20	IX, *D*2	H	Face, head; of animal (Hertz's list)
36	1	3	IX, *D*9	H	Tree
37*	52	41	X, *D*1	B	Spider, crab, lobster, crustacean
38	54	39	X, *D*1	H	Spider, crab, octopus, crustacean
39	33	25	X, *D*1	R	Crab, octopus
40	58	46	X, *D*1	K	Many-legged animal, or items 37, 38, 39
41*	17	14	X, *D*2	B	Dog
42‖	21	24	X, *D*4		Worm, caterpillar, serpent (*F* or *FC*)
43*	26	34	X, *D*4	H	Seahorse, or item 42
44	18	20	X, *D*4	K	Elongated animal, with use of color
45*	21	29	X, *D*5	BH R	Rabbit head
46	24	33	X, *D*5	K	Animal head with ears or horns, or item 45
47*	9	11	X, *D*12	B	Sheep

* Items remaining in the maximum saturation scale.

** Each group consists of 100 cases.

† Under "Location," numbers refer to Beck's D notation.

‡ Capital letters refer to Beck, Herts, Klopfer, and Rapaport, respectively.

§ Parentheses indicate overlapping items.

¶ Item 16 not correct as to Herts; noted too late to change.

‖ Item 42 belongs to no author; inserted to test with or without color.

Of the 47 items considered, 18 yielded a saturation of .63 in the experimental group. This was the maximum saturation possible, using no more than one of any group of "overlapping" items. These 18 items are designated with asterisks in Table 2.

Since a maximizing procedure was employed, the true characteristics of the *P* scale so developed should be estimated from Group B. Here the saturation is .51 and the reliability as estimated by Kuder-Richardson Formula 20 is .54. This was the highest internal consistency measure found for any of these scales, but obviously leaves much to be desired.

Table 3

Intercorrelations of *P* Scales ($N = 200$)

Author	*n*	Hertz	Klopfer	Rapaport	Saturation	*M*	σ
Beck	21	.88	.71	.86	.88	7.06	2.62
Hertz	20		.76	.82	.83	7.12	2.22
Klopfer	10			.68	.73	4.85	1.61
Rapaport	13				.71	5.80	1.81
Saturation	18					5.61	2.73

The intercorrelations of the five *P* scales for the entire group of 200 cases are given in Table 3. All have a great deal in common, as shown by the intercorrelations, none of which is lower than .68. High intercorrelations are to be expected because of overlap in rationale of the items and in actual item content. The saturation scale correlated .88 with Beck's, with which it has much content in common. Identical content makes all these intercorrelations high. It is of interest that in all but one of the items included in the saturation scale the determinant was form only. This may have some bearing on the point of view put forth by some authorities that determinants other than form should sometimes be considered in the selections of populars. In this connection, the question sometimes occurs as to whether the responses included as a popular should be defined quite specifically or in rather generic terms. For example, in Card 1 for the whole card, should one accept merely the bat, the butterfly, and the moth, but not a bird or a winged creature as a generic term? There are a number of points for which such a question could well arise. This point was tested during the investigation by replacing more specific terms by the more generic ones. In no instance was the saturation affected by more than a point or two.

Conclusions

Of the lists of populars proposed by four authors, Beck's scale exhibits the highest internal consistency. Its reliability, however, is definitely lower than that of acceptable psychometric devices. A revised list of 18 items, developed during the present investigation, is a slight improvement over Beck.

The present study indicates that orthodox psychometric methods can be used with projective material and will aid in the improvement of posited variables. The authors of this paper, a practicing clinician and a worker with conventional tests, were both surprised by the outcome. The internal consistency of the *P* scale, as developed by saturation analysis, was neither as high as the clinician expected nor as low as the psychometrician believed.

Variables which have a high degree of internal consistency may or may not be useful. Variables which have neither internal consistency nor external validity are certainly useless. It would appear that further examination of Rorschach variables is indicated.

Received February 22, 1954.

References

1. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* (2nd Ed., Rev.) New York: Grune & Stratton, 1949.
2. Beck, S. J., *et al.* The normal personality as projected in the Rorschach test. *J. Psychol.,* 1950, 30, 241-298.
3. DuBois, P. H., Loevinger, Jane, & Gleser, Goldine C. The construction of homogeneous keys for a biographical inventory. *Hum. Resour. Res. Cent., Res. Bull.,* 1952, No. 52-18.
4. Hertz, Marguerite R. *Frequency tables for scoring responses to the Rorschach inkblot test.* (3rd Ed.) Cleveland: Western Reserve Univer. Press, 1951.
5. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
6. Loevinger, Jane, Gleser, Goldine C., & DuBois, P. H. Maximizing the discriminating power of a multiple score test. *Psychometrika,* 1953, 18, 309-317.
7. Rapaport, D. *Diagnostic psychological testing.* Vol. II. Chicago: Year Book Publishers, 1946.
8. Rorschach, H., & Oberholzer, E. The application of the interpretation of form to psychoanalysis. *J. nerv. ment. Dis.,* 1924, 60, 225-248, 359-379.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Sex, Age, and Iconicity as Factors in Projective Film Tests[1]

Charles J. McIntyre
The Instructional Film Research Program[2]
The Pennsylvania State University

This study was undertaken to investigate the probable effectiveness of sound motion pictures as a medium for projective personality testing, and to attempt to define some of the characteristics which should be built into such a test.

Previous studies [2, 4, 5] have indicated that specially produced films may provide a powerful new tool for personality assessment. No attempt was made in these studies, however, to demonstrate that films might be superior to any conventional or widely used projective test, the Thematic Apperception Test (TAT), for example. Such a demonstration appears to be necessary if one is to justify the cost of producing a new test. Nor has any systematic attempt been made previously to indicate what characteristics such films should have in order to be maximally effective.

A related consideration involves the desirability of developing a criterion measure which not only would permit evaluating the limited variables tested in this study, but which could serve as a relatively quick and simple method for evaluating each of many stimulus characteristics that should be evaluated before constructing a motion picture projective test.

Since the success of picture projective tests appears to be dependent upon the degree to which subjects' (*S*s') interpretations of the stimuli are projections of their own personalities, the variables were evaluated in terms of the relative degree to which they permitted or encouraged projection onto the central figure (hero, protagonist).

The emphasis, which appears again and again in TAT literature, upon making the central figure like *S* with respect to age, sex, color, body-build, etc., in order to increase projection is directed to the end of making it possible for *S* to perceive the protagonist as being like himself. In fact, this appears to be the central meaning of projection in this context; namely, that an *S* has projected to the extent that his reported perception of the protagonist agrees with his perception of himself. This is the operational definition of projection used throughout this study.

The following hypotheses were tested.

1. There is a greater than chance agreement between an *S*'s perception of himself and his perception of an ambiguous film protagonist.

This hypothesis means, in terms of the operational definition, that persons viewing an ambiguous film will project their own self-perceptions onto the protagonist.

2. Provided that the total picture remains ambiguous, the amount of projection elicited by a protagonist will increase as the iconicity of the protagonist increases.

Iconicity is a term borrowed from semantics. It is used here to mean something akin to lifelikeness or realism. Apparently, the peculiar advantage of pictures, and especially motion pictures, over other forms of representational art is the great realism with which persons or things can be portrayed.

3. The *S*s will project more onto protagonists of the same sex than onto protagonists of the opposite sex.

4. The *S*s will project more onto protago-

[1] This paper represents the substance of a dissertation submitted in partia' fulfillment of the requirements for the Ph. D. degree at the Pennsylvania State University.

[2] This research was carried out under contract No. N6onr-269, T. O. VII between the Instructional Film Research Program, The Pennsylvania State University, and the Special Devices Center, Office of Naval Research.

nists of the same apparent age as the *S*s' than onto protagonists whose apparent age is different from the *S*s'.

Hypotheses 2, 3, 4 are all based upon the assumption that increased subject-protagonist similarity will result in increased projection. In hypothesis 2 the element of similarity is, of course, realism or lifelikeness. It says, in effect, that the probability for projection occurring is increased as the protagonists become more like real people and less like paintings or still pictures of people.

The assumption that increased similarity results in increased projection permeates the TAT literature. It was made specifically by Murray [7], Thompson [12], and Symonds [11]. It is implicit in the studies of Crandall [1], and Heppell and Raimy [6] when, in selecting pictures to elicit information in specific areas, they were careful to stipulate that at least one character in their pictures be of the same sex and age as the persons to be tested. In addition, among the rules set forth by Rotter for determining onto which figure an *S* is projecting, one finds, "The subject is likely to identify with a character of the same sex, either one of his own age or a previous age. For example, a young male adult might identify with a young male adult or a boy in the story but would be unlikely to identify himself with an older person, or a female of any age" [9, p. 82]. "Identify" in this context apparently means to project one's own characteristics onto a particular figure in the card.

It should be noted, however, that this assumption is not supported by the findings of Riess, Schwartz, and Cottingham [8, 10] and Weisskopf and Dunlevy [13, 14] which, while they do not indicate what conditions are necessary for eliciting maximum projection, do indicate that, whatever the necessary conditions may be, they are more subtle than simple physical similarity.

Procedures

The films. In all treatments, the stimulus materials were 16-mm. motion pictures which from the beginning of the fade-in to the end of the fade-out were each 35 seconds long. Five film versions were produced for each of five scenes, plus an additional film which was shown at the beginning of each experimental session as a practice exercise. Thus, there was a total of twenty-five different experimental films plus one practice film.

The five experimental scenes upon which the versions were built were selected from five TAT cards so as to embody the sex and age variables of the study. Since it was essential to duplicate as nearly as possible these scenes in the motion pictures, the selection of cards was limited, in that it was necessary to find persons who looked like the TAT characters they were portraying, and to find locations for shooting which were like those of the cards. An additional limiting factor was the necessity of selecting scenes in which the pose was relatively static rather than one implying action. The action versions of the scene had to conform to the general principle of ambiguity, and it was found in planning the scripts that any action coming before or after a dynamic still pose tended to define the meaning of the scene. When a pose was essentially static it was relatively easy to get an actor into or out of that pose without giving definition to the scene. With these limitations in mind, the following scenes were selected as the variables to test the age and sex hypotheses:

Scene 1. Variable: Young boy (TAT Card 1) A young boy is contemplating a violin which rests on a table in front of him.

Scene 2. Variable: Young man (TAT Card 20) The dimly illumined figure of a man (or woman) leaning against a lamppost in the dead of night.

Eron [3] in his normative study of the TAT found that 20 per cent of nonhospitalized *S*s misidentified the figure in card 20. To avoid contaminating the data because of misidentification, *S*s who were tested with this scene in the present study were required to indicate the sex and approximate age of the figure. Those who did not identify the figure as that of a young man were eliminated in the statistical analysis.

Scene 3. Variable: Young woman (TAT Card 8-GF) A young woman sits with her chin in her hand looking off into space.

Scene 4. Variable: Old man (TAT Card 7BM) A gray-haired man is looking at a younger man who is sullenly staring into space.

Although there are two persons in this scene, *S*s were told to make their judgments about the old man only.

Scene 5. Variable: Old woman (TAT Card 6BM) A short elderly woman stands with her back turned to a tall young man. The latter is looking downward with a perplexed expression.

As in Scene 4, there are two persons in this scene. The *S*s were told to make their judgments about the old woman only.

Scene 6. The practice scene shown to all groups (TAT Card 14). The silhouette of a young man (or woman) against a bright window. The rest of the picture is totally black.

The variable of iconicity was built into the five motion picture versions of each of the five scenes. Iconicity has been defined as lifelikeness or realism. In the development of the versions, an attempt was made to keep each successive version of a particular scene as nearly identical as possible, while adding elements that would make the characters more real or more lifelike than in the preceding version. Thus, Version 1, a photograph of a TAT card, was the least iconic since the figures in the card were merely drawn representations of people. The figures in Version 2 were more lifelike in that they were photographs of real people rather than drawings. In Version 3, the characters moved into the scene and thus were endowed with another attribute of real people, i.e., motion. Version 4 extended the principle of Version 3, that is, the characters manipulated and interacted with their environment. Version 5 attempted to add another dimension of behavior associated with people, namely, speech. Since the speech in Version 5 was made unintelligible in order to retain ambiguity, it may be that this version was actually less lifelike than the other versions rather than more, as intended. However, it can be seen that the intent in each case was to add increments of lifelikeness to each succeeding version in order to make the characters more like real, living people than in the preceding versions.

The measure of projection. For the purpose of this study projection has been operationally defined as the degree to which an individual's perception of a film protagonist agrees with his perception of himself. The Minnesota Multiphasic Personality Inventory (MMPI) was used as a pool from which items were drawn to sample *S*s' perception of the protagonist and their perception of themselves. That is, after the film sequence was shown, *S*s were asked to answer true or false to 26 questions which read, for example, "This is the sort of person who feels more intensely than most people do." After responding to these perception-of-protagonist questions, *S*s then took an MMPI for themselves during the course of which they would answer an item which read, "I think that I feel more intensely than most people do." When an *S*'s answer to the question with respect to the protagonist was the same as his answer with respect to himself, this was scored as a "unit of projection" and the sum of these was his "projection" score.

The full MMPI, rather than just the pertinent 26 items, was administered in order to obscure from *S*s the relationship between the items in the film test and the items in the inventory.

The items of the perception-of-protagonist test (the film test) were selected from the MMPI by two persons working independently. Only those items were used which both agreed met the following criteria:

1. The item should not be objectively answerable from the film. That is, there should be nothing about the appearance or action of the protagonist which would suggest a "correct" answer.
2. An item should not have a distinctly "abnormal" answer. Thus an item such as "Someone has control of my mind" although perhaps suitable according to criterion 1, would be unsuitable, since most people would answer this false for themselves and false for the protagonist merely by inference and thus, by the measure, be "projecting" with respect to this item.
3. In the interest of subject cooperation and motivation an item should have the appearance of being answerable from the film. An item such as, "My hands and feet are usually warm enough" would be rejected since an *S* would be likely to feel that he could not reasonably be expected to answer this item with respect to a film protagonist.
4. Items should be drawn from various scales of the personality inventory in approximately equal number.

The numbers of the 26 MMPI items used with each of the sequences were: 8, 13, 26, 32, 39, 54, 64, 79, 82, 91, 93, 94, 102, 165, 171, 242, 257, 267, 292, 299, 336, 362, 371, 399, 407, 416.

The subjects. The *S*s were 425 college students in elementary courses in education, psychology, and sociology. They were assigned randomly to the treatments. Each *S* saw only one of the twenty-five different sequences. Hence, the total subject pool was divided randomly into 25 groups of 7 men and 10 women each.

Test administration. The screen image was held constant at 58½ inches wide and 44 inches high. The arbitrary image size was dictated by the size of the smallest room (and therefore minimum projector-to-screen distance) available for the experiment.

In summary, the *S*s were told that it is well accepted that one of the chief characteristics of motion pictures is their ability to portray complex human emotions, feelings, and interactions, but that the ability of the actors is only one aspect of this. This experiment, they were told, was being conducted to find out about another aspect, namely, the ability of the audience to understand and interpret what they had seen. Hence, the *S*s were told they would see a very brief scene and would then be asked to make certain judgment about it. They were urged to guess when they could not be sure of their answers.

Following the instructions, the practice scene was shown. The *S*s then had 3 minutes to answer 20 questions written by *E* so as to be relatively easy to answer in order to encourage *S*s to think that the task was not really difficult. The same scene was then shown again and the *S*s given a short period to complete answering the questions. The *S*s did not know that this was a practice exercise. Following this the experimental film was shown, *S*s were given 3 minutes to answer the 26 selected items, the same film was shown again, and *S*s were given another brief period to complete answering the questions for that film. After the test papers were collected, it was explained that *E* was interested in relating their judgments of the film characters to certain aspects of personality and that, therefore, they were being asked to complete a personality inventory. The *S*s were permitted to work at their own speed and leave when they had finished.

Results

The experiment has the form of a 5 × 5 × 2 factorial design. There were five scenes, each varying five ways (versions) with the *S*s grouped by sex. Analysis of variance was the principal statistical technique. The data were tested, and it was shown that the conditions of normally distributed scores and homogeneity of variance, necessary in the analysis of variance, were met.

Table 1

Analysis of Variance of Projection Scores

Source of variation	*df*	Mean square	*F*
Between versions	4	57.3705	1.34
Between scenes	4	341.3176	7.98**
Between sexes	1	2.7880	—
Version × scene	16	54.1073	1.65
Version × sex	4	25.7330	—
Scene × sex	4	54.7306	1.67
Version × scene × sex	16	32.7392	1.77*
Within groups	375	18.5386	
Total	424		

* Indicates significance at 5% level.
** Indicates significance at 1% level.

The analysis of variance is summarized in Table 1.

It will be seen in Table 1 that the differences between the means of the scenes were highly significant. It was therefore appropriate to determine the nature and extent of these differences by means of *t* tests. These results are summarized in Table 2.

Table 2

Differences between Means of Scenes

Mean	of Scene	Mean	of Scene	Difference
12.86	1	11.78	2	1.08
12.86	1	10.93	3	1.93**
12.86	1	16.19	4	—3.33**
12.86	1	13.31	5	— .45
11.78	2	10.93	3	.85
11.78	2	16.19	4	—4.41**
11.78	2	13.31	5	—1.53*
10.93	3	16.19	4	—5.26**
10.93	3	13.31	5	—2.38**
16.19	4	13.31	5	2.88**

* Indicates significance at 5% level.
** Indicates significance at the 1% level.

It will be seen in Table 2 that the mean projection score for Scene 4 (old man) was significantly greater than any other scene, and the mean score for Scene 5 (old woman) was significantly greater than the mean of Scenes 2 and 3.

Because of the nature of the measure of projection the need for additional analysis was indicated. Since an individual's projection score was the sum of the times he answered an item the same way on each of two tests, merely by chance an *S* might be said to be "projecting" with respect to a particular item 50 per cent of the time and thus might be expected to get a "projection" score of 13 (26/2) by chance alone. The mean projection score over all *S*s was 13.01 Because the obtained and chance mean scores were virtually identical, chi-square values were computed for the agreement between answers to each item in both tests. The frequency of obtained chi squares for each of the 26 items was compared to the chance distribu-

tion of chi square. The resulting chi-square value yielded a p of approximately .15. Hence, the observed distribution of chi square for the 26 items did not differ significantly from the distribution that would be expected to arise if only chance were operating.

The implications of these findings in terms of the specific hypotheses are discussed below.

1. Hypothesis 1 predicted that *S*s would project their own characteristics onto an ambiguous film protagonist. The last finding reported, i.e., that the frequency of agreement between the same items on the two tests (which was the measure of projection) did not differ significantly, seems to indicate, if considered alone, that projection as defined and measured in this experiment did not occur. On the other hand, the F ratio for differences between scenes was significant well beyond the 1 per cent level.

It could be true, of course, that the present experiment was the "less-than-one-time-out-of-a-hundred" in which the differences occurred by chance. By definition, this event is unlikely. A more likely interpretation of this apparent contradiction is that no measurable degree of projection occurred with respect to Scenes 1, 2, and 3 but that *S*s did project their own characteristics onto the protagonists of Scene 4 and, possibly, Scene 5. However, while there was sufficient nonchance agreement between the film test and MMPI items to raise the mean projection score on Scene 4 significantly above all other scenes, and on Scene 5 above all but one other scene, these nonchance agreements, when grouped with the apparently chance scores of the other scenes, were insufficient to produce chi-square values which differed significantly from chance.

It, therefore, can be said that *S*s were found to project onto some ambiguous film protagonists, but the result did not indicate that the phenomenon was general.

2. Hypothesis 2 predicted that projection would increase with the iconicity of the protagonist. The iconicity variable was embodied in the versions. The between-versions F of 1.34 in Table 1 indicated that no significant differences between the means of versions were found. In view of the fact that the scores of at least three of the five scenes did not differ significantly from chance, it was to be expected that the versions would not differ. When Scenes 4 and 5, in which there appears to have been projection not attributable to chance, were analyzed apart from the other scenes, no significant differences between the means of the versions of those scenes were found. Hypothesis 2 was not confirmed.

3. The third hypothesis predicted that *S*s would project more onto protagonists of the same sex than onto protagonists of the opposite sex. In Table 1, it will be seen that the scene × sex interaction term was not significant. In fact, both men and women projected most onto the protagonist of Scene 4, an old man. Hence, this hypothesis was not confirmed.

4. Hypothesis 4 predicted that *S*s would project most onto protagonists of the same apparent age as themselves. Hence the greatest projection should have occurred onto Scene 2 (young man) and Scene 3 (young woman). It will be seen in Table 2 that these two scenes were the least effective in eliciting projection. The most effective scenes were 4 and 5 in which the protagonists were an old man and old woman respectively. Hypothesis 4 was not confirmed.

Discussion

This study was undertaken primarily to obtain information about the use of motion pictures for projective testing. While the results do not permit valid conclusions to be drawn about the relative effectiveness of films per se, some subjective impressions of the experimenter might appropriately be discussed here.

One of the severest handicaps that a film suffers, when used as in this experiment, lies in the short time which may be given to *S* to view the stimulus scene. (In this instance, two showings of 35 seconds each.) In the conventional TAT administration, *S* may take virtually as much time as he wishes for looking at the card. Furthermore, he has the card before him while he is telling his story. It may well be that this makes an important contribution to the value of a still picture for eliciting projection because, with an essentially unlimited time to study the picture, *S* is provided with an always present source of material from which to elaborate his fantasies.

It is rather difficult to provide as much time

in a motion picture as an *S* might effectively use, while retaining the necessary element of ambiguity and adding the presumably important element of realism. A motion picture is realistic partly because it depicts people behaving—acting and interacting. Normally, the more a person is shown behaving, the more the situation is defined, i. e., it loses ambiguity. While motion pictures can be produced showing people behaving without defining the situation, it is not certain, since it is not one's usual experience with motion pictures or with life, that such a procedure does not detract from the realism of the scene. A film loop device which would permit the repeated showing of a single short sequence could be used but, again, it is not certain that this would not impair the realism of the scene inasmuch as people normally do not repeat an identical action over and over. The same difficulty applies to sound, for it would seem that an ambiguous sound heard repeatedly either ceases to be ambiguous, or convinces the subject that it was not intended to be interpretable.

These considerations are subjective and should not preclude further research in the area. They do, however, indicate some of the problems which must be met if a motion picture test similar to the TAT is to be developed.

Conclusions

1. The perception of an ambiguous protagonist as being like oneself does not appear to be as generalized a phenomenon as some of the TAT literature would indicate. Such projection was found to occur but to a very limited extent.

2. The experiment did not yield empirical evidence to determine the probable effectiveness of motion pictures as a medium for projective personality testing. That is, projection was not found to vary with the iconicity of the stimuli. There is subjective evidence, however, to suggest that a film test following the TAT paradigm may not be the most effective use of the motion picture medium for this purpose.

3. With respect to sex and age, this study did not support the hypothesis that *S*s will project more onto protagonists who are physically similar to themselves. Indeed, this study seemed to support the conclusion that such similarity may actually discourage projection.

Summary

A study was undertaken to investigate the probable effectiveness of sound motion pictures as a medium for projective personality testing, and to attempt to define some of the characteristics which should be built into such a test.

Short motion pictures, based upon five TAT cards whose central figures differed by sex and age, were produced in such a way as to incorporate varying degrees of realism.

The projection associated with each of the three variables was measured and compared. Projection was operationally defined as the extent to which a subject's perception of himself agreed with his perception of the protagonist.

Little projection was found to occur. The projection that did occur was associated with the two parental-figure protagonists. No differences in projection associated with the realism variable were found.

Received December 11, 1953.

References

1. Crandall, V. J. Induced frustration and punishment-reward expectancy in TAT stories. *J. consult. Psychol.*, 1951, **15**, 400-404.
2. Eiserer, P. E. The relative effectiveness of motion and still pictures as stimuli for eliciting fantasy stories about adolescent-parent relationships. *Genet. Psychol. Monogr.*, 1949, **39**, 205-278.
3. Eron, L. D. A normative study of the TAT. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
4. Fulchignoni, E., Sbordoni, Gigliola, & Tarroni, Evelina. Il film e i metodi proiettivi (The film and the projective methods). *Arch. Psicol. Neurol. Psichiat.*, 1950, **11**, 97-169.
5. Gemelli, A. Le film comme méthode projective. *Acta Psychol.*, 1950, **7**, 190-195.
6. Heppell, H. K., & Raimy, V. C. Projective pictures as interview devices. *J. consult. Psychol.*, 1951, **15**, 405-411.
7. Murray, H. A. *Manual for the Thematic Apperception Test.* Cambridge: Harvard Univer. Press, 1943.
8. Riess, B. F., Schwartz, E. K., & Cottingham, Alice. An experimental critique of assumptions underlying the Negro version of the TAT. *J. abnorm. soc. Psychol.*, 1950, **45**, 700-709.
9. Rotter, J. B. Thematic apperception tests: suggestions for adminstration and interpretation. *J. Pers.*, 1946, **15**, 70-92.
10. Schwartz, E. K., Riess, B. F., & Cottingham, Alice. Further critical evaluation of the Negro version of the TAT. *J. proj. Tech.*, 1951, **15**, 394-400.

11. Symonds, P. M. Criteria for the selection of pictures for the investigation of adolescent fantasies. *J. abnorm. soc. Psychol.*, 1939, **34**, 271-274.
12. Thompson, C. E. The Thompson modification of the TAT. *Rorschach Res. Exch.*, 1949, **13**, 469-478.
13. Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379-390.
14. Weisskopf, Edith A., & Dunlevy, G. P. Body similarity between subject and central figure in the TAT as an influence on projection. *J. abnorm. soc. Psychol.*, 1952, **47**, 441-445.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

A Note on the Use of a Paper-and-Pencil Form of the MMPI *Hs* Scale for Hospital Use [1, 2, 3]

Sol Charen
Catholic University of America

The use of a single scale of the MMPI, according to Machover and Anderson [1], is often desirable in clinical practice. This is especially true in hospital clinics and outpatient departments where some method of screening patients with neurotic symptoms from those with organic problems is needed.

The 33 items of the *Hs* scale of the MMPI were mimeographed on a single sheet in the same sequence in which they appear in the group form booklet of the MMPI. Subjects were 50 hospitalized tuberculosis patients shortly to be discharged as arrested cases. Their median age was 35 years with the age range from 21 to 69 years.

The method of testing was identical to that used by Machover and Anderson except that the time interval between test and retest was one week. The group-form booklet of the MMPI was used and testing was done in groups. Half the patients were given the paper-and-pencil form of the *Hs* scale and a week later were retested with the booklet. The other 25 were tested in reverse order.

The mean *T* score of the 50 subjects on the experimental form was 56.14 against a mean of 57.87 for the group booklet form of the scale with p equal to .73 for a two-tailed test. Mean *T* scores for all methods of testing ranged from 55.20 to 59.80. The Pearson product-moment correlation between group booklet form of the *Hs* scale and the paper-and-pencil form was .980.

The paper-and-pencil form of the *Hs* scale of the MMPI gave almost identical *T* scores as a retest with the group booklet form. The correlation between the two was .980. The paper-and-pencil form of the *Hs* scale seems, then, worth while for use in hospital clinics and outpatient departments where quick testing and screening of patients is desirable.

Brief Report
Received June 16, 1954.

[1] From the Veterans Administration Hospital, Baltimore, Maryland. The statements and conclusions reached in this article are those of the author and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2] The author wishes to thank Dr. Ross McLean, Chief of Medicine, VA Hospital, Baltimore, Maryland for his help in making available the patients used in this research.

[3] An extended report of this study may be obtained without charge from Sol Charen, 200 Rhode Island Avenue, N. E., Washington 2, D. C. or for a fee from the American Documentation Institute. To obtain it from the latter source order Document No. 4320 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Reference

1. Machover, S., & Anderson, Helen J. Validity of a paper-and-pencil form of the MMPI Psychopathic Deviate scale. *J. consult. Psychol.*, 1953, 17, 459-461.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Transcendence Indices, Extent of Personality Factors in Fantasy Responses, and the Ambiguity of TAT Cards[1]

Douglas T. Kenny
University of British Columbia

This paper is the third in a series concerned with testing the widely accepted assumption of a monotonic relationship between stimulus ambiguity and extent of personality factors in the evoked fantasy. Any study exploring this relationship must concern itself with two primary methodological considerations, namely, the specifications of the meanings assigned to the concepts of both *stimulus ambiguity* and *personality content* and the way in which these terms are to be measured.

Weisskopf [6, 7] has presented evidence showing that pictures with low physical ambiguity elicit less projection than more highly structured ones. Physical "impoverishment" of the stimulus, by incomplete line tracings and varying exposure time, was used as an indicator of stimulus ambiguity. Projection was defined in terms of a transcendence index which is based on the mean number of responses "which go beyond pure description." In an earlier paper it was pointed out by Bijou and Kenny that such findings " . . . may be used to evaluate the proposition under consideration provided one assumes that (*a*) stimulus ambiguity is properly defined in terms of physical properties and, (*b*) there is a direct relationship between amount of fantasy and amount of personality factors introduced" [1, p. 204].

Employing a different set of measurement operations, Kenny and Bijou [2] found that with increasing psychological ambiguity there is an initial increase in the extent of personality factors revealed and then a decrease. In contrast to Weisskopf's procedure, ambiguity was defined in terms of psychological rather than physical properties. Ambiguity was defined in terms of the estimated number of interpretations that judges thought would apply to each TAT picture. In analyzing the significance of personality factors based on TAT stories, Stephenson's [4] *Q*-array technique was utilized rather than unidimensional scales or transcendence indices.

It seems pertinent to relate these two different approaches that have been used in testing the monotonic relationship postulated between stimulus ambiguity and extent of personality factors revealed in fantasy responses. The present research aimed at investigating the relationship between projection, as defined by transcendence indices, and extent of personality factors revealed to TAT pictures, as determined by a *Q*-array technique. It seems worth while to test Weisskopf's assumption that there is a direct relationship between her transcendence indices and the extent of personality factors introduced into TAT protocols. As yet, there is no empirical evidence to indicate that transcendence indices can be used as indicators of the amount of clinically interpretable material elicited by a picture. In addition, an attempt was made to see how transcendence indices varied as a function of physical ambiguity, psychological ambiguity, and stimulus ambiguity involving a combination of both physical and psychological properties.

Method

Ascertaining Stimulus Ambiguity

The three procedures for determining stimulus ambiguity were:

Psychological ambiguity. From the first study [1], 15 TAT cards varying in three degrees of psycho-

[1]The author wishes to express his appreciation to Dr. Edith A. Weisskopf-Joelson for providing him with a transcendence index scoring manual. He is indebted to Professor E. S. W. Belyea for photographing and making the projector slides of the TAT pictures.

logical ambiguity were obtained. A group of 51 judges ranked 21 TAT cards in terms of the estimated number of interpretations each card possessed. By this procedure three sets of five cards, varying in three degrees of increasing psychological ambiguity, were selected. Cards 5, 15, 18BM, 11, and 19 were classified as the most ambiguous sets. Cards 4, 7BM, 8BM, 12M, and 13MF made up the intermediate group. Cards 12BG, 1, 2, 9BM, and 17BM comprised the least ambiguous set.

Physical ambiguity. In order to induce physical ambiguity, changes in the exposure time of the TAT cards were used. For the present study, the 15 TAT cards were photographed with approximately the same intensity as the standard set. The cards, on individual slides, were projected onto a screen, and the subjects were simply asked to describe the projected picture. The pictures were shown to 80 subjects, using three to eight subjects at one time. All subjects were drawn from introductory psychology classes at the University of British Columbia. Forty male subjects described the pictures under a 5-second exposure time (high physical ambiguity), and another 40 male subjects described the pictures under a 2-minute exposure time (low physical ambiguity).

Combined physical and psychological ambiguity. The 15 TAT cards were classified in terms of both physical and psychological dimensions of ambiguity by making use of Murray's placement of the cards. It will be recalled that Murray divided the cards into two groups of ten, cards numbered 10 and below being more structured than those numbered 11 and above. Murray's placement of the cards was based on the physical characteristics of the cards as well as on diagnostic value. Of the 15 TAT cards, seven are from Murray's first 10 and eight are from his second ten.

Measurement of Personality Factors

As indicated, the aim of this research was to determine the relationship between two measures of fantasy responses to TAT pictures, a Q-array technique and transcendence indices. Transcendence indices for each of the 15 TAT cards were obtained by scoring the descriptions that the subjects gave to the cards under the two exposure times. Weisskopf's manual on the scoring for transcendence indices was employed. Separate transcendence indices were worked out for the different exposure times. For a detailed description of the Q-array scoring method and the administration method of the TAT pictures, the reader is referred to the second paper in this series [2]. Suffice it to say here that the 15 TAT cards were administered individually to 18 male college students. Two well-trained clinical psychologists, utilizing the Q-array method, judged the significance of the personality factors "injected" into the TAT stories. The correlation of the scores between the two clinicians was satisfactory ($r = .68$).

Results

The most important question likely to arise is: What, if any, is the relationship between the ranking of the TAT cards for revealingness of the personality as determined by the Q-array method and the ranking of the same cards in terms of transcendence indices? The rank-order correlations between personality revealingness and the transcendence indices were .62 and .64 for the 5-second and the 2-minute exposure groups respectively. Statistical significance is beyond the .01 level for the ρ of .64 and at the .05 level for the ρ of .62. The data clearly support Weisskopf's contention that transcendence indices are related to personality revealingness in TAT pictures.

A second question asked is: To what extent does the transcendence index vary as a function of the physical ambiguity of the stimulus? Table 1 presents this finding in terms of the difference between exposure times. The data confirm Weisskopf's findings in showing that subjects reliably increase their transcendence indices under long exposure times.

Table 1

Analysis of Variance of Transcendence Indices

Source of variation	*df*	Mean square	*F*
Between exposure times	1	1643.26	44.94**
Subjects in same group	78	36.56	
Total between subjects	79		
Psychological ambiguity level	2	73.86	18.37**
Ambiguity × exposure time	2	17.58	4.37*
Pooled subjects × ambiguity	156	4.02	
Total within subjects	160		
Total	239		

* Significant at 5 per cent level.
** Significant at 1 per cent level.

A third question asked is: Do transcendence indices vary with degree of psychological ambiguity when the transcendence scores are classified under three sets of pictures varying in three degrees of increasing psychological ambiguity? The F for Between Psychological Ambiguity Level, as shown in Table 1, indicates significant differences between the three levels of psychological ambiguity. To determine which sets of cards show the differences,

t tests were made between the sets using the interaction of pooled subjects × ambiguities as the error estimate. This analysis showed that the intermediate ambiguity set had significantly higher transcendence indices than either the most or least ambiguous sets. The differences were significant at the 1 per cent level of confidence. There were, however, no differences in transcendence indices between the most ambiguous and least ambiguous cards. This finding, combined with the previous finding [2] that the cards in the median ambiguous set reveal significantly greater personality factors than those in the least and most ambiguous sets, shows that those cards which have higher transcendence indices are also the cards which reveal the most about personality.

Data in Table 1 also show a significant interaction between length of exposure time and psychological ambiguity level. Examination of the differences involved in this interaction shows that the intermediate ambiguous set yielded higher transcendence indices than the most ambiguous set under long exposure time, but not under short exposure time; also, the intermediate set showed significantly higher transcendence indices than the least ambiguous set, under either short or long exposure time.

The final question asked is: Are there higher transcendence indices for the TAT cards from the last ten than from the first ten? Analysis of variance results showed that there were no differences in the transcendence indices between these sets of cards. Hence, Murray's hypothesis of a difference in degree of projection between these two sets of cards is not supported by the data. The interaction between this twofold classification of ambiguity and time of exposure was not significant.

Discussion

Two of the most interesting findings which emerge from this study are that the extent of personality material revealed in a picture is correlated with the transcendence index of the picture and that cards from the intermediate psychologically ambiguous set yield higher transcendence indices than either the least or most ambiguous sets. The former finding supports Weisskopf's assumption of a direct relationship between amount of transcendence and amount of personality factors. The latter result may be accounted for in terms of three possible factors which are described in greater detail elsewhere [2]. Briefly, transcendence indices may not have increased with increasing psychological ambiguity because (*a*) anxiety, which may mount with increases in ambiguity, inhibited greater projections; (*b*) the cue patterns in the three sets of cards do not vary as assumed; or (*c*) there may have been qualitative differences among the three sets of cards. Any three of these plausible explanations may account for the obtained results.

The results of the present study, as well as Weisskopf's research, do not conform to the theoretical expectation that increases in physical ambiguity should be accompanied by increases in transcendence indices. In terms of the previous formulation of this problem [2], it may be hypothesized that a decreasing number of cue constellations available for selective reactions produces response tendencies which do not contain much in the way of individual projections.

Contrary toMurray's hypothesis, cards from his more structured series did not yield higher transcendence indices than cards from the less structured series. This finding substantiates the previous conclusion [2] that Murray's most ambiguous set does not elicit more significant personality factors than his least ambiguous set. On the other hand, Weisskopf [5] found that the well-structured series actually elicited higher transcendence indices than the least structured series. This discrepant finding between the two studies may be explained in terms of the differences in the TAT cards used in the two studies.

Summary

This paper describes an experiment relating the concepts of transcendence and extent of personality factors in elicited fantasy. The results are consistent with Weisskopf's assumption that the extent of personality material revealed in a TAT picture is correlated with the transcendence index of the picture. In addition, it was shown that with increasing psychological ambiguity there is an initial increase in transcendence and then a decrease. The theoretical expectation that transcendence indices

should increase with increases in physical ambiguity was not substantiated. Finally, the expectation that TAT cards from Murray's second series should elicit higher transcendence indices than cards from his first series was not upheld.

Received January 30, 1954.

References

1. Bijou, S. W., & Kenny, D. T. The ambiguity values of TAT cards. *J. consult. Psychol.*, 1951, **15**, 203-209.
2. Kenny, D. T., & Bijou, S. W. Ambiguity of pictures and extent of personality factors in fantasy responses. *J. consult. Psychol.*, 1953, **17**, 283-288.
3. Murray, H. A. *Thematic Apperception Test.* Cambridge, Mass.: Harvard Univer. Press, 1943.
4. Stephenson, W. *The study of behavior.* Chicago: Univer. of Chicago Press, 1953.
5. Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379-390.
6. Weisskopf, Edith A. An experimental study of the effect of brightness and ambiguity on projection in the Thematic Apperception Test. *J. Psychol.*, 1950, **29**, 407-416.
7. Weisskopf-Joelson, Edith A., & Lynn, D. B. The effect of variations in ambiguity on projection in the Children's Apperception Test. *J. consult. Psychol.*, 1953, **17**, 67-70.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

A Psychotherapy Case Showing Progress but No Decrease in the Discomfort-Relief Quotient

Edward J. Murray, Frank Auld, Jr., and Alice M. White[1]
Yale University

The purpose of this paper is to demonstrate that a case may show progress even though there is no drop in verbal tension. Verbal tension was measured by the Discomfort-Relief Quotient (DRQ), an objective method of determining the tension expressed in the sentences of a patient during psychotherapy interviews. The DRQ was developed by Dollard and Mowrer [4] to measure tension during social casework interviews. Mowrer [12, 13] and Kauffman and Raimy [10] have gone further with it and have concluded that a tension drop during psychotherapy, as measured by the DRQ, indicates therapeutic progress.

The DRQ was compared with a content-analysis procedure which shows changes that can be interpreted as indicating progress. The method of content analysis was developed by the senior author [14] on the basis of concepts of motivation and conflict derived from learning theory and psychoanalytic theory.

The DRQ and the content analysis were applied to the case of Mrs. B., a psychotherapy case which has received a detailed clinical study by Dollard, Auld, and White [2]. The existence of this qualitative study of the case will enable the interested reader to evaluate better the clinical significance of the quantitative results presented here.

The DRQ Study of the Case

The procedure for scoring the DRQ has been described in detail by Dollard and Mowrer [4]. Briefly, each unit is judged as indicating *discomfort* (drive, want, tension, pain, suffering), *relief* (drive reduction, reward, pleasure, relaxation), or *neither*. The DRQ is the number of discomfort units, divided by the number of discomfort units plus relief units. The resulting quotient is multiplied by 100 to provide a scale ranging from 0 (no tension) to 100 (all tension).

In applying the DRQ to the case of Mrs. B., two graduate students independently scored summaries of all 17 therapy hours. These students were psychologically sophisticated but unfamiliar with the case. They scored the summaries in a random order so that they would not be influenced in their scoring by the "plot" of the case or by preconceptions about the direction of tension change. Their agreement was high ($r = +.92$, $df = 15$, $p < .01$). Figure 1 shows the average DRQ of the two scorers for each of the therapy hours. It can be seen that, although there is some fluctuation in the DRQ from hour to hour, there is no consistent trend throughout the course of therapy. There is neither a gradual decline in ver-

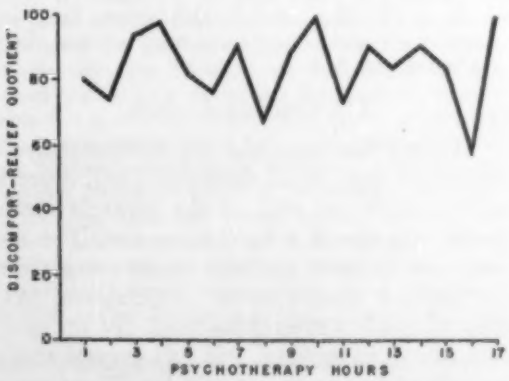

Fig. 1. The Discomfort-Relief Quotient based on summaries of 17 consecutive therapy hours with Mrs. B.

[1]Mr. Murray's share of this work was supported by a U. S. Public Health Service research fellowship. Part of the study was also supported by a grant from the U. S. Public Health Service to Professor John Dollard, who sponsored this research. The authors wish to thank Drs. Dollard, George F. Mahl, and Neal E. Miller for a critical reading of the manuscript.

bal tension nor a sudden drop near the end of treatment.

It is possible that the DRQ showed no decline because the summaries did not truly reflect tension expressed in the patient's sentences. In order to explore this possibility, two other psychologically sophisticated people applied the DRQ to the verbatim typescripts of six of the interviews. These scorers were given the pages of the six interviews in random order. The interviews were selected for typing in the following way: five successive blocks of three or four hours were delineated and one hour was selected randomly from each block. Hour 11 was added because it had been previously typed for the Dollard, Auld, and White book. The agreement of the two scorers was high ($r = +.96$, $df = 4$, $p < .01$). The averaged results are shown in Figure 2. Again, no trend is apparent.

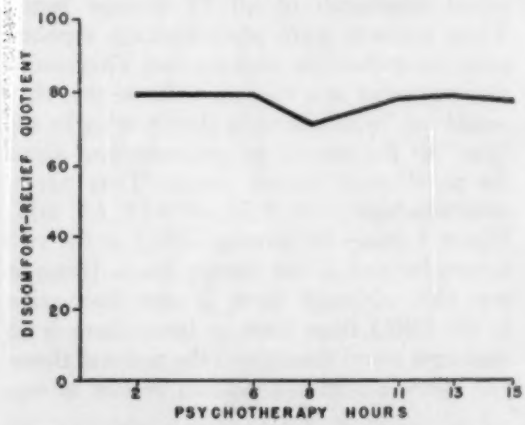

Fig. 2. The Discomfort-Relief Quotient based on verbatim typescripts of six randomly selected therapy hours with Mrs. B.

The Content Analysis

Murray has described his content-analysis method in more detail elsewhere [14]. Scorers in this study put each of the patient's statements into one of several categories. The results are reported for four major categories: (*a*) hostility toward mother and children, (*b*) general hostility toward husband, (*c*) sex conflict related to husband, and (*d*) general anxiety related to no one in particular. These categories include both approach and avoidance tendencies in each area. Thus, hostility includes both expression of hostility and expression of anxiety or guilt about hostility. The sex conflict with the husband includes sexual desire, sexual anxiety, and hostility specifically related to sexual frustration.

The main judge in this study scored the typescripts of six therapy hours in a random order. A psychologically naive scorer was used to check reliability. This person scored two hours with the main judge for training purposes and then scored the remaining four hours independently. Agreement between the two scorers for the four hours was high: for hostility to mother and children, $r = +.997$, $df = 2$, $p < .01$; for general hostility to husband, $r = +.995$, $df = 2$, $p < .01$; for sex conflict with husband, $r = +.993$, $df = 2$, $p < .01$; and for general anxiety, $r = +.956$, $df = 2$, $p < .05$. In order to control for differences in the total number of statements from hour to hour, the raw frequency scores were converted into percentages of the total number of statements in the hour.

The sum of all four categories reveals no striking change throughout therapy, as shown in Figure 3. Whatever slight trend exists can be accounted for almost entirely by the high point on Hour 6. Although the precise shape of this curve differs from the DRQ curve in Figure 2, the over-all impression is the same.

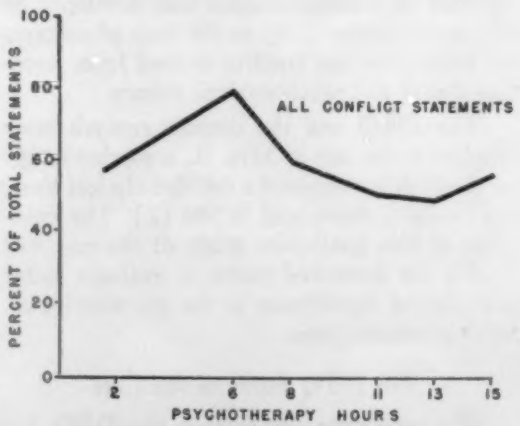

Fig. 3. Conflict statements of four major categories in six randomly selected therapy hours with Mrs. B.

The content analysis shows important changes in the patient's speech when the separate areas of conflict are considered. During the course of therapy, the patient shifted from

talking about her conflict with her mother and children to talking about her conflicts with her husband. This is shown in Figure 4, where "hostility conflict related to mother and children" and the sum of the conflict categories relating to the husband (i.e., both general hostility and sex) are plotted.

The nature of the conflicts about the husband also changed during therapy. Figure 5 shows the category of general hostility to the husband and the category of sex conflict with the husband. The general hostility conflict was replaced by the sex conflict during the course of therapy. The general hostility conflict consisted of criticisms of the husband's attitudes toward social life, vacations, etc., and some guilt about these criticisms. The sex conflict included statements about the patient's own frustration, statements blaming her husband, and eventually statements showing some insight into her own inhibitions.

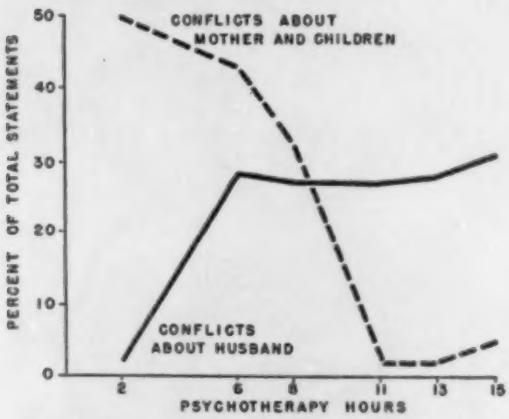

Fig. 4. Conflict statements about mother and children, and conflict statements about husband, in six randomly selected therapy hours with Mrs. B.

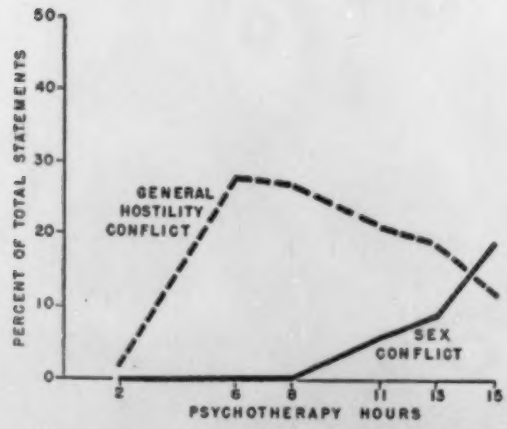

Fig. 5. General hostility conflict statements about husband, and sex conflict statements about husband, in six randomly selected therapy hours with Mrs. B.

The content analysis, then, shows a change during the course of therapy from talking about a hostility conflict about the patient's mother and children to talking about a hostility conflict about unimportant matters concerning her husband and, finally, to talking about a fundamental sexual conflict about her husband.

Discussion

According to Dollard, Auld, and White [2], Mrs. B. made some progress in therapy, although her case was not a "finished" one. Their account of the case is briefly as follows:

> Mrs. B. began therapy with a defensive "cover story." She complained that her mother had stolen away her children and that she (the patient) was unable to stand up for her rights in the situation. According to Dollard, Auld, and White, this baby-snatching was not a primary problem; indeed, Mrs. B. had very little trouble asserting herself toward her mother or anyone else. The therapist refused to accept the cover story as an adequate explanation. He also insisted that she could not avoid discussing her marital life. Mrs. B. began to complain about her husband. At first she complained about unimportant matters such as his lack of interest in parties. The therapist found it difficult to believe that these things had the overwhelming importance ascribed to them. He noted that the patient had said nothing at all about her sex life. Mrs. B. gradually began to talk about her sex life. She said at first that it was satisfactory; she failed to see any connection between sex and her presenting symptoms of depression. But as she followed the rule of free association she began talking about her sexual frustration and her husband's inadequacy. Later she saw the role her own inhibitions played in the situation. She finally realized that her sexual feelings toward her father had been transferred to her husband along with her horror of committing incest. She had unconsciously rejected her marriage and her responsibilities to her children as well. Toward the end of therapy the patient reported that her husband had spontaneously commented that she "seemed much better in bed — more interested now." Undoubtedly Mrs. B. had not solved all of her problems, but she had gained some insight.

The results of the independent content anal-

ysis are in remarkable agreement with the qualitative analysis. The content analysis shows a decrease in the category of "hostility conflict about mother and children," which corresponds to what Dollard, Auld, and White call the "cover story." The content analysis also shows the shift in verbalizations first to a general hostility conflict about the husband and then to a sexual conflict about the husband. The advantage of the content analysis is that it describes precisely and quantitatively therapeutic events which the qualitative analysis describes in an impressionistic way. Moreover, the content analysis has been demonstrated to be reliable. The agreement between the content analysis and the qualitative analysis shows that the content analysis measures meaningful and theoretically relevant things.

The DRQ did not show the decrease which is said to accompany progress. Several other investigations have also found no relationship between the DRQ and therapeutic progress. Hunt [7, 8] and Hunt and Kogan [9] found a correlation of only .20 between DRQ decrement and caseworkers' ratings of movement. The failure of the DRQ to drop in the present case was due to the fact that the patient verbally expressed tension about something or other throughout therapy. The progress in the case was due to an increase in insight which occurred when the verbalizations shifted toward a basic conflict. The patient learned something new about herself which apparently affected her behavior outside of therapy.

An important problem is the relationship of the DRQ to the tension actually experienced by the patient. Tension may have remained at the same level throughout therapy with the patient gradually becoming aware of its origin in her sexual conflict. In this case the DRQ would have accurately reflected the tension experienced by the patient. However, there is reason to doubt such an accurate reflection since the DRQ does not correlate very highly with physiological measures of tension [13]. It is much more likely that Mrs. B. experienced more tension as she began discussing her sexual conflicts than she had while elaborating her cover story. In this connection, it is perhaps significant that the final content category of general, or free-floating, anxiety consistently increased during therapy.

There are many examples in the literature of the need to increase tension in order to get therapeutic results. Freud [6] found it was necessary to break through *la belle indifférence* of the hysteric to achieve therapeutic results. The isolation of emotion from thought in the obsessive-compulsive has proved a stumbling block in therapy; it is only when such a patient *feels* tense that progress is made [5]. Alexander and French [1] have cited many cases in which the frequency of contacts was purposely manipulated to increase tension. The authors believe that ". . . every analysis should be conducted on as high an emotional level as the patient's ego can stand without diminishing its capacity for insight" [1, p. 30]. Miller's theory of conflict [11] provides a rationale for these empirical findings. According to this theory, when a therapeutic tactic enables a patient to approach closer to a feared goal, the absolute level of fear is increased. The result has been called the negative therapeutic effect [3], since the increased fear often brings about a return of symptoms. Therefore, although we would expect an eventual reduction of tension in a successful patient's everyday life, we would not necessarily expect a reduction of tension in the therapeutic situation, particularly in short-term therapy.

The foregoing analysis raises the problem of why the patient stays in therapy in the face of increasing tension. The answer is partly that many patients leave therapy when anxiety increases. In successful cases, the tendency to leave seems to be counterbalanced by several little understood factors. First, there are many rewarding features of the transference relationship [3]. Second, the symptoms and frustrated drives which originally motivated the patient to seek help continue to operate. Third, the solution of minor problems eases the tension in the patient's daily life to some extent and reinforces hope of solving the more important problems.

Summary and Conclusions

1. The purpose of this study was to show that a psychotherapy case in which there was no drop in verbal tension might show progress in other ways. The Discomfort-Relief Quotient was used as a measure of verbal tension,

and a content analysis in terms of motivation and conflict was used to show other verbal changes.

2. Two separate applications of the DRQ to the case of Mrs. B. showed no change in verbal tension during therapy. Each application was demonstrated to be reliable.

3. The content analysis showed that during therapy the patient progressed from a discussion of a "cover story" about a hostility conflict with her mother to a discussion of conflicts related to her husband. Furthermore, she progressed from discussing a general hostility conflict about relatively unimportant matters in relation to her husband to discussing much more fundamental sexual conflicts. The content analysis was shown to be reliable.

4. The results of the content analysis were shown to be consistent with an independent, qualitative clinical analysis.

5. It was pointed out that the DRQ may not reflect the tension experienced by a patient. It was also pointed out that although an eventual tension reduction in a successful patient's everyday life would be expected, a reduction of tension in the therapeutic situation would not necessarily be expected.

6. We conclude that although the DRQ may be a good measure of the tension expressed in a patient's sentences, a study of this one case provides no support for the view that the DRQ is useful for assessing therapeutic progress. In this case, other measures of the content of the patient's speech proved more illuminating than the DRQ.

Received February 8, 1954.

References

1. Alexander, F., & French, T. M. *Psychoanalytic therapy.* New York: Ronald, 1946.
2. Dollard, J., Auld, F., & White, Alice M. *Steps in psychotherapy.* New York: Macmillan, 1953.
3. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
4. Dollard, J., & Mowrer, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.,* 1947, **42,** 3-32.
5. Freud, S. *Collected papers.* Vol. II. London: Hogarth, 1924.
6. Freud, S. *Collected papers.* Vol. IV. London: Hogarth, 1925.
7. Hunt, J. McV. Measuring the effects of social casework. *Trans. N. Y. Acad. Sci., Ser. II,* 1947, **9,** 78-88.
8. Hunt, J. McV. A social agency as a setting for research — the Institute of Welfare Research. *J. consult. Psychol.,* 1949, **13,** 69-81.
9. Hunt, J. McV., & Kogan, L. S. *Measuring results in social casework.* New York: Family Service Association of America, 1950.
10. Kauffman, P. E., & Raimy, V. C. Two methods of assessing therapeutic progress. *J. abnorm. soc. Psychol.,* 1949, **44,** 379-385.
11. Miller, N. E. Comments on theoretical models, illustrated by the development of a theory of conflict behavior. *J. Pers.,* 1951, **20,** 82-100.
12. Mowrer, O. H., Hunt, J. McV., & Kogan, L. S. Further studies utilizing the Discomfort-Relief Quotient. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research.* New York: Ronald, 1953. Pp. 257-295.
13. Mowrer, O. H., Light, B. H., Luria, A., & Zeleny, M. P. Tension changes during psychotherapy with special reference to resistance. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research.* New York: Ronald, 1953. Pp. 546-640.
14. Murray, E. J. A case study in a behavioral analysis of psychotherapy. *J. abnorm. soc. Psychol.,* 1954, **49,** 305-310.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

The Effect of Childhood Diseases on IQ Variability[1]

Lillian Wagenheim
The University of Southern California

Two previous studies by the writer showed a relationship between the early contraction of measles, chicken pox, mumps, and German measles and later school achievement. This relationship was found among boys who had contracted these diseases under age three and most significantly in those at age two.

In the present study, the records of 493 boys were used. The data obtained included IQ's, achievement grade level in reading and arithmetic, and the ages of contraction of the following childhood diseases: measles, chicken pox, German measles, mumps, whooping cough, diphtheria, scarlet fever, tonsillitis, pneumonia.

Cases were considered variable if two IQ scores resulting from tests administered during different years showed a difference of 10 or more points.

The contraction of whooping cough, tonsillitis, chicken pox, and measles under age five was related at the 1% or 3% level to the variability in IQ. The contraction of mumps, German measles, and pneumonia under age five did not show a significant relationship to IQ variability.

In order to discover the effect of childhood diseases in general on IQ variability, the cases were regrouped for the earliest contraction of any disease.

Boys who had contracted any disease under age five were more frequently found in the variable group than in the nonvariable group. Chi square was 11.65 and significant beyond the 1% level of confidence.

An especially interesting and relevant finding occurred when the records of children of different school achievement levels were grouped separately. The cases had been divided on the basis of achievement in reading and arithmetic for a previous study. A child who functioned at least .7 of a grade below his chronological age level was considered an underachiever.

Results significant at the 1% or 2% level were found in the separate achievement groups at different age levels. The children who were underachievers in both reading and arithmetic had the lowest median IQ. They also showed more variability in IQ when childhood diseases had been contracted at ages two through four. Children who were underachievers in *either* reading *or* arithmetic were average in IQ, and their intelligence was more variable when diseases had been contracted at ages one through three. Those who were achieving in both reading and arithmetic were superior in median IQ and showed more variability in intelligence when the diseases were contracted at ages zero through two years.

It seems premature to attempt to answer the question as to whether the disturbance has a direct relationship to encephalomyelitis following childhood diseases or whether the trauma to the nervous system interferes with the psychosocial (ego) development of the child. Apparently there is a hierarchy of ages at which the diseases are most deleterious and which appears to coincide with the emergence of language ideation at different levels of intelligence.

Brief Report
Received June 21, 1954.

[1] An extended report of this study may be obtained without charge from Lillian Wagenheim, University of Southern California, Los Angeles 7, California or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4319 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

The Relationship Between Motor and Cognitive Inhibition

Julian Meltzoff[1]
VA Regional Office, Brooklyn, N. Y.
and Murray Levine
Philadelphia VA Mental Hygiene Clinic

In psychoanalytic theory, the gradual shift from the primary to the secondary process involves the delay and control of impulse expression and the substitution of symbolization, fantasy, and planning for immediate response on a behavioral level. Rapaport's [5] psychoanalytic theory of thinking points to the inhibition of action designed to gratify immediate needs as a primary condition for the development of fantasy. Young children are characteristically lacking in the ability to inhibit motor activity and tend to verbalize whatever thoughts come to mind. Piaget [3] and Luria [1] have illustrated the use of speech and symbolic circuition for the control of behavior. This is paralleled in the Rorschach test by a gradual emergence of human movement (*M*) responses with maturation [9]. Certain psychotic groups also display faulty inhibition of thoughts and motor behavior, and are similarly lacking in the ability to produce *M*.

Rorschach [6] implied a triadic relationship between motor activity, motion perception, and the cognitive processes. Experiments by Singer, Meltzoff *et al.* [2, 7, 8], testing hypotheses deduced from Werner and Wapner's [11] sensory-tonic theory, have demonstrated one aspect of that triadic relationship, namely the vicarious relationship between motor inhibition and illusory motion perception. The purpose of the present investigation is to study the linkage between motor inhibition and the inhibition of a cognitive process.

[1]The authors are greatly indebted to VA trainees Louis Pietroforte, Norman Slamecka, and Norris Weinberg for their assistance in this experiment, and to Jerome and Dorothy Singer for their stimulating criticisms and suggestions.

The Experiment

Hypothesis. It is hypothesized that there is a direct relationship between motor inhibition and the inhibition of an association process. More specifically, it was predicted that *S*s with greater ability to voluntarily inhibit motor activity would be able to inhibit learned associations and produce new ones quicker than those less adept at inhibiting motor activity.

Procedure. Eighty volunteer male and female *S*s were selected haphazardly from a university population. They were requested to memorize a verbally presented list of ten pairs of words, to a criterion of one perfect reproduction. Instructions were those used in the Wechsler Memory Scale [10]. The list of pairs was composed of stimulus words and the most common associations to them, as reported by Rapaport [4]. Association time was recorded to the nearest half second.[2] After the list was memorized, *S*s were instructed to respond, upon presentation of the stimulus word, with any word *other than* the learned paired associate. Responses were timed in the same manner as previously. The *S*s were then instructed to write the phrase "New Jersey Chamber of Commerce," as slowly as possible, within the boundaries of parallel lines drawn one-half inch apart. They were reminded at 45-second intervals to write as slowly as possible.

The motor inhibition task of enforced slow writing had been used successfully in prior experiments [2]. As far as the cognitive inhibi-

[2]In the experimental procedure, the dependent variable was measured prior to the independent variable, so that any errors in timing were not biased errors.

tion task is concerned, it was reasoned that the production of new associations to words with already memorized associates would involve two processes: (*a*) inhibiting the learned associate and (*b*) finding and reporting a different association. We shall refer to the latter process as *associative facility*. There was no apparent reason to believe it to be correlated with the ability to inhibit motor activity. The assumption was therefore made that associative facility would vary randomly among the experimental population, and that any obtained differences would be primarily due to systematic variation in the ability to inhibit associations. A partial test of this assumption was subsequently carried out, and is reported in the discussion of results. The *Ss*' introspections about the cognitive inhibition task were solicited, and they uniformly reported either a conscious attempt to suppress an image of the learned associate, or, in the absence of an image, an awareness that there was a word to be avoided before an association could be produced.

Treatment of the data. The population of 80 *Ss* was divided into long and short motor inhibition time (MIT) groups at the median of the distribution of obtained inhibition times. It is of interest to note the extreme differences among *Ss* in their capacity to inhibit motor activity. Time for writing the experimental phrase varied all the way from 20 seconds to 44 minutes, 36 seconds. The significance of the difference between the means of these two groups of 40 *Ss* on cognitive inhibition time (CIT) was calculated by Fisher's *t*. Analysis of covariance was carried out to take into consideration differences between the reaction times of the groups on the initial criterion trial.

Results

As can be seen in Table 1, the mean CIT of the poor motor inhibitors (i.e., those with short MIT) is appreciably greater than that of the good motor inhibitors (i.e., those with long MIT). This difference is statistically significant at $p < .005$, with a t of 3.38. Analysis of covariance reveals that this difference cannot be accounted for by the initial differences in reaction time that exist between the two groups. The obtained F value is significant at $p<.005$.

There was relatively little overlapping between the two groups. Seventy-seven per cent of *Ss* who were least able to inhibit motor activity took longer to inhibit associations than the mean time of the better motor inhibitors. On the other hand, only 12 per cent of the better motor inhibitors took longer than the mean time of the short MIT group to inhibit associations.

Table 1

Mean Reaction Time and Cognitive Inhibition Time of Long and Short Motor Inhibition Time Groups

Measure	Short MIT group	Long MIT group
Mean reaction time, Paired associates*	1.0	0.9
Mean CIT*	2.9	2.0

* Times are group means of individual Ss' average reaction times for ten responses.

Table 2

Analysis of Covariance of Reaction Time and Cognitive Inhibition Time of Long and Short Motor Inhibition Time Groups

Source	*df*	Adjusted mean square	*F*
Between groups	1	1278.680	9.94
Within groups	77	128.455	

Discussion

In view of the fact that the experimental sample was divided at the median, instead of representing extremes of the population, the clear difference between the groups lends strong confirmation to the hypothesis. The ability voluntarily to inhibit motor activity apparently goes hand in hand with the ability to inhibit associations as measured by the experimental task. This finding, however, is dependent upon the correctness of the assumption underlying the choice of the particular task selected to measure thinking inhibition. The primary issue is the relegation of associative facility to a secondary role while considering inhibition as the primary determinant of association times in the experimental task.

In order to provide an experimental check on this reasoning, the cognitive inhibition task was repeated with 30 new *Ss* from the same general population. These *Ss* were also required to

give a single association to each of ten different words taken from the same source as the original pairs. The response time was taken as a measure of associative facility. A random half of the *S*s had the pairs first, while the remainder began with the single-word associations. Division of the population into fast and slow CIT groups reveals a mean association time of 1.3 seconds for the rapid thought inhibitors and a mean of 1.6 seconds for the slow group. This difference is in the inconclusive range, with a t of 1.72 ($p = .10$). There is a positive, but not statistically significant, correlation (rho $= .32$) between the two variables.

These results suggest that associative facility is perhaps more of a factor than was originally assumed, but nevertheless a secondary one which could not reasonably account for very much of the highly significant difference between the groups in the original experiment. It is quite plausible that the ability to produce new associations rapidly is *one* of the means by which people are able to accomplish the work of inhibiting thoughts. Those who have this ability are to this extent better equipped to inhibit. There are probably other abilities as well that facilitate inhibition.

The findings of this investigation support the speculation and evidence from diverse sources relevant to the development of the secondary process cited in the introduction. The results also bear, at least indirectly, upon our understanding of Rorschach *M* responses. Werner's sensory-tonic theory suggests a vicarious relationship between perceptual and motor activity and enables one to predict that a decrease in motor activity will be accompanied by an increase in the perception of illusory motion. Experiments by Singer, Meltzoff *et al.* [2, 7, 8] confirm hypotheses deduced from this theory and show that the voluntary inhibition of motor activity results in an increase in *M* production. They also have demonstrated that those who are more productive of *M* responses are characteristically better able to inhibit motor activity. The present study adds the additional information that those who are more capable of inhibiting motor activity are also better able to inhibit established associations and produce new ones. It does not necessarily follow that *M* is a measure of the capacity to inhibit associations as well as motor activity, but these accumulated data would lead us to hypothesize that such a relationship exists.

It is of interest to report the application of these experimental techniques to the study of the psychopathology of an individual clinical case. The patient was an intelligent manic whose behavior was characterized by motor and verbal hyperactivity and flights of ideas. He was examined while under treatment with the drug Tolserol, a muscular relaxant. He displayed no atypical motor behavior during the course of the examination. He gave neither *M* nor color responses on the Rorschach, could inhibit his motor activity for only 29 seconds (the mean of the total experimental population is 5 minutes, 48 seconds),[3] and was very poor at inhibiting associations with a mean time of 6.4 seconds (the mean of the total experimental population was 2.4 seconds).[3] It was concluded that despite the superficial effects of Tolserol, the patient was basically unable to inhibit either motor activity or thoughts. About two weeks later, medication was discontinued, and he had a violent episode marked by uninhibited verbalizations and extreme hyperactivity necessitating his being placed in seclusion.

Summary and Conclusions

This study was designed to test the hypothesis of a direct relationship between motor inhibition and the inhibition of a cognitive process. Eighty *S*s were divided at the median into two groups on the basis of motor inhibition time, as measured by an enforced slow writing task. The difference between these two groups in cognitive inhibition time, as measured by *S*s' speed in inhibiting learned word association pairs and producing new associations to the stimulus words, was found to be statistically significant in the predicted direction.

It is concluded that *S*s with greater ability to voluntarily inhibit motor activity are able to inhibit learned associations and produce new ones more quickly than those less adept at inhibiting motor activity.

Received January 18, 1954.

[3]These means are merely illustrative. They are *not* norms for clinical use.

References

1. Luria, A. R. *The nature of human conflicts.* New York: Liveright, 1932.
2. Meltzoff, J., Singer, J. L., & Korchin, S. J., Motor inhibition and Rorschach movement responses: a test of the sensory-tonic theory. *J. Pers.,* 1953, **21,** 400-410.
3. Piaget, J. *The language and thought of the child.* New York: Harcourt, Brace, 1932.
4. Rapaport, D. *Diagnostic psychological testing.* Vol. 2. Chicago: Year Book Publishers, 1946.
5. Rapaport, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.
6. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
7. Singer, J. L., Meltzoff, J., & Goldman, G. D. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.,* 1952, **16,** 359-364.
8. Singer, J. L. & Spohn, H. E. Some behavioral correlates of Rorschach's experience-type. *J. consult Psychol.,* 1954, **18,** 1-9.
9. Thetford, W., Molish, H. B., & Beck, S. J. Developmental aspects of personality structure in normal children. *J. proj. Tech.,* 1951, **15,** 58-78.
10. Wechsler, D. A standardized memory scale for clinical use. *J. Psychol.,* 1945, **19,** 87-95.
11. Werner, H., & Wapner, S. Sensory-tonic field theory of perception. *J. Pers.,* 1949, **18,** 88-107.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Classification of "Suicidal" Patients[1]

Albert Rosen, William M. Hales,[2] and Werner Simon
VA Hospital, Minneapolis, Minnesota and University of Minnesota

The term "suicidal" has been used to describe patients demonstrating heterogeneous kinds of behavior, such as (*a*) suicide thoughts, (*b*) suicide threats, (*c*) suicide attempts, and (*d*) severe depression. Some investigators have grouped together psychological test data from such individuals, as well as from those who later committed suicide, in an attempt to develop and evaluate psychometric devices for the detection of "suicidal" patients [2, 3, 4, 5, 6]. It is important to consider whether or not such grouping of patients under the class term "suicidal" is too broad for effective use of psychological tests in the study and detection of the patient who is apt to attempt suicide.

Only one published study is relevant to the problem of differentiating "suicidal" subgroups by means of psychological tests. Farberow [1] analyzed MMPI records and other tests of three groups of patients in a California veterans psychiatric hospital. Each group contained 32 male patients as follows: (*a*) patients who made a suicide attempt within a relatively short period before admission, (*b*) patients who made suicide threats, and (*c*) patients selected at random, with no history of suicide attempts, threats, or thoughts (Control Group). Tests were administered to each patient in the first group *after* the attempt was made. The most striking finding in the MMPI data was the marked discrepancy in elevation of mean MMPI scale scores between the Suicide Attempt and Suicide Threat Groups. The scores for the Suicide Threat Group indicated a much greater severity of disturbance. Moreover, the mean scores for the Suicide Attempt and the Control Groups were not significantly different from each other.

Farberow's results from the use of other tests also emphasized that patients who have made a suicide attempt are considerably dissimilar from those expressing suicide threats. The purpose of this paper is to report an investigation conducted in the Psychiatry Section of the Minneapolis Veterans Administration Hospital. This study consisted essentially of a replication of that portion of Farberow's research employing the MMPI in the evaluation of "suicidal" patients.

Selection and Description of Subjects

From the summaries of all discharged male patients, two groups were selected, as follows: (*a*) patients who had made a suicide attempt shortly before admission to the hospital (Suicide Attempt Group), and (*b*) patients who expressed suicide thoughts (Suicide Thought Group). Although finer grouping of the cases was desired, it was not possible from the available file data. The MMPI for each of these patients was then obtained, but was included in the study only if it had been administered within 15 days after admission and prior to electroshock or insulin therapy. A few additional cases were acquired from more recent discharge summaries in order to obtain samples of 50 Suicide Attempt cases and of 100 Suicide Thought cases.

A third sample consisting of a cross section of the psychiatric patient population was included in the study for comparison with the Suicide Attempt and the Suicide Thought Groups. In a previous unpublished study, a sample of 250 psychiatric patients-in-general had been selected. This group consisted essentially of every third patient admitted during

[1] From the psychiatric service, VA Hospital, Minneapolis, Minnesota, and the Divisions of Psychiatry and Clinical Psychology of the University of Minnesota.

[2] W. M. Hales is now at the Area Medical Office, Veterans Adminstration, Fort Snelling, St. Paul, Minnesota.

a two-year period who had taken the MMPI. For the purposes of the present study, 39 cases were removed from this patient-in-general group because their discharge summaries indicated one or more attempts at, or thoughts of, suicide. These 211 cases will be referred to as the Control Group.

For each group, data were obtained on diagnosis, age, education, marital status, and intelligence. Only summary statistics will be presented because the major interest of this investigation is in the comparison of personality test scores. Primary diagnoses of the Suicide Attempt, the Suicide Thought, and the Control Groups are listed in Table 1. The groups do not differ significantly in broad diagnostic categories. However, among patients classified as neurotic, those who have made suicide attempts or who have engaged in suicide thinking are much more frequently diagnosed as depressive reaction than are patients-in-general. It is quite possible that the diagnosis of depressive reaction is sometimes influenced by the evidence of suicidal behavior or mentation.

Biographical data for the Suicide Attempt, Suicide Thought, and Control Groups, respectively, are as follows: median years of age—32, 32, 32; median years of education—11, 12, 10; percentage of group who are married or widowed—68, 52, 58; percentage who are divorced or separated—8, 18, 12; percentage in group below 110 IQ as roughly estimated from the Shipley-Hartford test—26, 38, 44.

Table 1

Primary Diagnoses of Patients in the Suicide Attempt, Suicide Thought, and Control Groups

Diagnosis	Per cent of Attempt Group ($N = 50$)	Per cent of Thought Group ($N = 100$)	Per cent of Control Group ($N = 211$)
Depressive reaction	34	22	4
Anxiety reaction	16	24	28
Other neurotic	6	7	26
Total neurotic	56	53	58
Paranoid schizophrenia	16	14	10
Involutional melancholia	0	7	4
Other psychotic	4	8	8
Total psychotic	20	29	22
Behavior disorder	18	15	15
Organic	4	3	3
No diagnosis	2	0	2
Total sample	100	100	100

The median number of days between hospital admission and MMPI testing for the Suicide Attempt, Suicide Thought, and Control Groups was 3, 4, and 3, respectively.

In the Suicide Attempt Group, the number of patients making each of the various modes of attempt is as follows: Sedation—18; slash-

Table 2

MMPI Scale Raw Scores for Suicide Thought, Suicide Attempt, and Control Groups

Scale	Mean raw score			Variance raw score			*F* ratio
	Thought	Attempt	Control	Thought	Attempt	Control	
L	3.63	3.28	3.76	6.37	6.68	4.36	.88
F	9.63	8.68	7.40	40.83	39.62	27.72	5.28**
K	12.15	12.64	13.73	25.41	36.87	30.04	3.07*
Hs	14.37	11.14	13.76	48.73	75.04	49.77	3.63*
Hs + .5 *K*	20.68	17.74	20.88	41.86	45.47	42.47	4.77**
D	33.08	28.14	28.60	40.71	44.28	49.93	16.16**
Hy	29.76	26.66	28.47	41.76	53.70	43.58	3.64*
Pd	23.28	23.04	20.92	36.38	50.44	37.54	5.83**
Pd + .4 *K*	28.24	28.26	26.40	32.26	32.27	32.60	4.58*
Mf	26.93	26.08	25.33	28.51	29.83	29.60	2.99
Pa	14.40	13.46	12.31	19.90	24.89	14.73	8.66**
Pt	26.97	22.70	21.38	97.99	151.89	115.74	9.11**
Pt + 1.0 *K*	39.10	35.34	35.12	57.89	70.54	60.13	9.15**
Sc	26.74	22.40	20.22	180.43	222.80	150.78	8.47**
Sc + 1.0 *K*	38.87	35.04	33.95	128.63	117.92	92.31	7.73**
Ma	18.20	19.58	17.76	23.80	33.56	28.33	2.41
Ma + .2 *K*	20.58	22.14	20.47	19.50	25.16	23.15	2.57
Si	37.82	32.00	32.10	118.09	134.48	120.18	9.72**

* Significant at the .05 level.
** Significant at the .01 level.

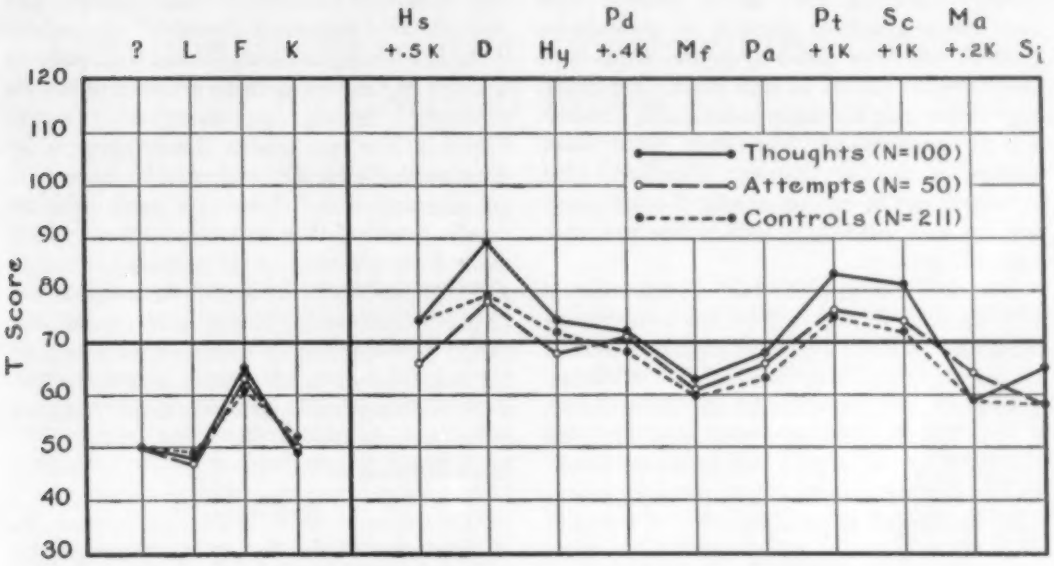

Fig. 1. Mean MMPI *T*-score profiles for Suicide Thought, Suicide Attempt, and Control Groups.

ing wrists—8; gas or carbon monoxide—7; shooting—4; hanging or strangulation—3; jumping from bridge—2; jumping in path of auto—2. For 6 cases the method was not indicated. For 2 cases listed only as "slashing," one also used strangulation, and another gas.

MMPI Results

All calculations were made with raw scores, and the raw score mean and variance of each MMPI scale for the Suicide Attempt, Suicide Thought, and Control Groups are presented in Table 2 for reference purposes. The Cannot Say score was zero in most cases, and is therefore not considered further. Mean *T* scores based on the raw scores are shown in Figure 1. In order to determine the significance of the differences between means, an analysis of variance test was made for each scale. The Bartlett test indicated no significant difference in the variances for any scale. The *F* ratio for each analysis of variance is listed in the last column of Table 2, from which it can be seen that almost every scale differentiates in some manner between the groups. For each scale with a significant *F* ratio the significance of the difference between the means of each pair of the three groups was computed by the *t* test. Only differences found to be significant at or beyond the .05 level are listed in Table 3.

Table 3

t Ratios for Significant Differences in Mean MMPI Scores between Suicide Thought, Suicide Attempt, and Control Groups

Scale	Thought > Attempt	Thought > Control	Attempt > Control	Control > Attempt	Control > Thought
F		3.22			
K					2.38
Hs	2.62			2.34	
Hs + .5 *K*	2.59			3.04	
D	4.16	5.38			
Hy	2.67				
Pd		3.10	2.15		
Pd + .4 *K*		2.65	2.07		
Pa		4.09			
Pt	2.01	4.26			
Pt + 1.0 *K*	2.77	4.18			
Sc		4.11			
Sc + 1.0 *K*		3.92			
Si	2.69	4.26			

Note.—t ratios of 1.96, 2.58, and 3.29 are significant at the .05, .01, and .001 levels, respectively.

On every scale in Table 3, except for *K*, the Suicide Thought Group is significantly higher in mean score than either the Suicide Attempt or the Control Group. The Suicide Attempt Group is significantly higher than the Control Group on the *Pd* and *Pd* + .4 *K* scales, and significantly lower on *Hs* and *Hs* + .5 *K*. These findings indicate that, as a group, patients with thoughts of suicide are much more

severely disturbed than either patients who have made a suicide attempt or patients-in-general, and that patients in the latter two categories are similar to each other. One might hypothesize that for many individuals, a reduction in the degree of psychiatric disturbance follows a suicide attempt. However, this hypothesis can be evaluated only if one has test data for such individuals both before and after a suicide attempt.

The results of this study are quite consistent with those of Farberow [1]. Both studies indicate that patients who have attempted suicide are markedly different from those who threaten or contemplate suicide. In any investigation of the "suicidal" patient, it appears probable that the most meaningful results would be obtained from an analysis of groups which are as finely delineated as possible.

Summary

Psychological test data from patients demonstrating various kinds of "suicidal" behavior are usually grouped together by investigators who are attempting to develop suicide detection devices. Farberow has shown that there are marked differences between groups of patients who have made suicide attempts and those who have threatened suicide. The present paper reports a replication of a portion of Farberow's research. The MMPI records of three groups of male, veteran psychiatric patients were analyzed. The groups consisted of (*a*) patients who made a suicide attempt shortly before admission (Suicide Attempt Group), (*b*) patients who expressed thoughts of suicide (Suicide Thought Group), and (*c*) patients selected at random, with no evidence of suicide attempts, threats, or thoughts (Control Group). The two studies closely agree in indicating that groups of patients who express suicide mentation and threats are much more severely disturbed than either patients who have made a suicide attempt or patients-in-general, and that patients in the latter two categories are similar to each other. It may be concluded that in any investigation of suicide, most meaningful results might be obtained if data for individuals with various kinds of "suicidal" behavior are differentially classified.

Received December 7, 1953.

References

1. Farberow, N. L. Personality patterns of suicidal mental hospital patients. *Genet. Psychol. Monogr.*, 1950, **42**, 3-79.
2. Fisher, S. The value of the Rorschach for detecting suicidal trends. *J. proj. Tech.*, 1951, **15**, 250-254.
3. Hertz, Marguerite R. Suicidal configurations in Rorschach records. *Rorsch. Res. Exch. & J. proj. Tech.*, 1948 **12**, 3-58.
4. Hertz, Marguerite R. Further study of "suicidal" configurations in Rorschach records. *Rorsch. Res. Exch. & J. proj. Tech.*, 1949, **13**, 44-73.
5. Simon, W., & Hales, W. M. Note on a suicide key in the Minnesota Multiphasic Personality Inventory. *Amer. J. Psychiat.*, 1949, **106**, 222-223.
6. White, Mary A., & Schreiber, Hanna. Diagnosing "suicidal risks" on the Rorschach. *Psychiat. Quart. Suppl.*, 1952, **26**, 161-189.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Duration of Illness as a Prognostic Indicator in Mental Illness

C. H. Swensen, Jr.[1] and G. R. Pascal[1]
University of Tennessee

There is unanimous agreement among professional workers in the field of psychiatry that the sooner treatment begins after the development of mental illness, the greater the chances for recovery. As Kalinowsky and Hoch put it, "The most important factor in evaluating prognosis today is the duration of illness prior to treatment; complete agreement exists among all workers that there is an inverse relationship between the duration of illness and recovery" [1, p. 84].

It was the purpose of this study to evaluate the prognostic significance of duration of illness by studying its relationship to outcome of illness after several other variables, also significantly related to outcome, had been controlled.

Method

The data used in this study were obtained from 486 cases admitted to Western Psychiatric Institute and Clinic (WPIC), University of Pittsburgh, between November 1942 and July 1950. These cases included all persons admitted who met the following criteria: age between 15 and 55, no known brain damage, no mental deficiency, no serious physical illness, and no previous attacks of mental illness. This sample is more fully described in a previous paper [2].

The patients were judged to be either "improved" or "unimproved" as of one year following their discharge from WPIC. Patients were judged on the basis of social workers' or psychiatrists' reports. They were considered improved if they were out and working on a level comparable to that on which they had worked prior to becoming ill, and they were judged unimproved if still in a hospital, or if they were at home being cared for.

Duration was measured in months; the time that elapsed between the first time the patient or relatives sought medical help for the patient and the time he was admitted to the hospital for treatment was considered the duration. It is admitted that this allows for a considerable period of time for the patient to have been ill before it was realized by his relatives, particularly in the case of those whose illnesses developed insidiously, but this criterion was used because it was felt to be the most reliable for the purpose of this study.

The relationship between outcome of illness and 21 variables was determined. These 21 variables were: sex, age, marital status, nationality, occupation, economic level, body type, education, intellectual status, affective expression, orientation, direction of aggression, diagnosis, age at onset of illness, duration of illness, length of hospitalization, status upon discharge, precipitating stress, treatment, religion, and type of onset. It was found that 11 of these variables were related to outcome at the .01

Table 1

Summary Statistics of the Total Groups and Controlled Sample Group on Duration of Illness

Statistic	Total group	Controlled sample group
N	486	180
Mean months		
Improved	21.70	34.08
Unimproved	38.59	23.18
Standard deviation		
Improved	22.74	37.12
Unimproved	27.90	20.59
t†	7.31**	2.45*
r_{bis}‡	.38**	—.14

† Between improved and unimproved within group.
‡ Between duration and outcome.
* Significant at .05 level.
** Significant at .01 level.

[1] At the University of Pittsburgh when this research was done.

level of confidence. These 11 were: marital status, affective expression, orientation, direction of aggression, diagnosis, type of onset, duration of illness, hospital stay, status on discharge, precipitating stress, and treatment. The relationship between these variables and outcome is discussed more fully in a previous paper [2].

The correlation between duration of illness and outcome for this total sample of 486 was .38. This is presented in Table 1. It will be noted that the unimproved group had a mean duration of illness of 38.59 months, or 17 months more than the mean for the improved group.

The relationships between duration of illness and the other variables that were significantly related to outcome of illness are contained in Table 2. Duration was related to five of the variables (affective expression, direction of aggression, type of onset, precipitating stress, and treatment) at the .01 level. A long duration of illness tended to be associated with an inappropriate affect or an absence of affect, with aggressive impulses directed toward the environment, with the insidious onset, with relatively little precipitating stress, and the patients with a long duration tended to be treated more often with shock therapy than with other types of therapy.

Table 2

Relationship between Duration of Illness and Other Significant Variables

Variable	Chi square	*p*
Affective expression	26.9	.01
Orientation	4.2	.05
Direction of aggression	24.3	.01
Diagnosis	11.3	.05
Type of onset	79.8	.01
Precipitating stress	19.3	.01
Treatment	23.5	.01
Marital status	5.2	.30

After the relationships between these variables and outcome had been determined, the total group was divided into two groups, an improved group and an unimproved group. These groups were then matched for all of the significant variables except duration of illness. This matching was accomplished by selectively dropping cases from the two groups until there were no significant differences between the improved and unimproved groups on any of the variables other than duration that were significantly related to outcome of illness. There were 180 cases remaining, 90 improved and 90 unimproved, when this matching had been completed. The differences between the controlled sample groups on the significant variables are presented in Table 3. A check revealed that there were no differences between the groups on any of the variables not significantly related to outcome. After this matching was completed, the relationship between duration of illness and outcome of illness for the controlled groups was determined.

Table 3

Differences between the Controlled Sample Groups in the Significant Variables

Variable	Statistic*
Diagnosis	chi square = .00
Marital status	chi square = .10
Orientation	chi square = .51
Treatment	chi square = .00
Affect	$t = .64$
Direction of aggression	$t = .20$
Type of onset	$t = .89$
Hospital stay	$t = .08$
Precipitating stress	$t = .82$

* All below .10 level of confidence.

Results and Discussion

A comparison of the controlled sample groups revealed that the unimproved group had a lower mean duration of illness than the improved group, and that this difference was significant at the .05 level. The biserial correlation coefficient, which had been .38 for the total group of 486, was reduced to —.14. (These results are contained in Table 1, which was presented earlier.) These figures tend to give the impression that under the controlling procedure used in this study, a longer duration is a better sign than a short duration. However, a comparison of the frequencies having various durations for the improved and unimproved groups, as presented in Table 4, reveals that there is little difference between the unimproved and improved groups on number of patients falling into the various categories of duration. In the process of eliminating the effect of the other variables, the improved group was left with several patients with extremely

Table 4

Distribution of Controlled Sample Groups on Duration of Illness

Months duration	Improved group	Unimproved group
Under 6	12	12
6 – 11	12	17
12 – 23	18	11
24 – 35	13	20
Over 36	35	30
Chi square	4.40	

long duration of illness, thus raising the mean for the group.

The net result would seem to be that the elimination of the effect of the other variables significantly related to outcome also eliminated the relationship obtaining between duration and outcome. And it would appear that the importance of duration as a prognostic sign is derived chiefly from its relationship to other variables of more basic importance that are significantly related to outcome. It would appear that the safest conclusion that could be reached is that the mere passage of time is in itself neither good nor bad so far as the possibility of "curing" a mental illness is concerned, but rather it is the processes that are allowed to continue unchecked during this passage of time that are of importance to the outcome of a mental illness.

Summary

It was the purpose of this study to determine the relationship between duration of mental illness and the outcome of illness when the effect of other variables, significantly related to outcome, had been controlled.

A sample of 486 patients was used. The relationship between outcome of illness and 21 variables was determined. Then the group was divided into an improved and an unimproved group, and cases dropped from each of the groups until there were no significant differences between the groups on any of the variables other than duration of illness.

It was found that after this controlling was completed there were no significant differences between the groups on duration of illness. This result was interpreted as indicating that the prognostic importance of duration of illness derived from its relationship to other variables related to the basic disease process, that the mere passage of time presaged neither good nor ill for mental illness, but rather, that it was the processes allowed to progress unchecked during this period that were of prognostic importance in mental illness.

Received February 23, 1954.

References

1. Kalinowsky, L. B., & Hoch, P. *Shock treatment and other somatic procedures in psychiatry.* New York: Grune & Stratton, 1946.
2. Pascal, G. R., Swensen, C. H., Feldman, Dorothy A., Cole, Mary E., & Bayard, Jean. Prognostic criteria in the case histories of hospitalized mental patients. *J. consult. Psychol.*, 1953, **17**, 163-171.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Rigidity in Conceptual Thinking in Exogenous and Endogenous Mentally Retarded Children

J. G. McMurray
University of Toronto[1]

It is noted occasionally by clinicians that exogenous mentally retarded children, as compared with endogenous, are significantly less able to perform such memory-for-designs tests as that found at year IX on the Revised Stanford-Binet Intelligence Scale. At least one investigator [4] has suggested that their inferior performance is not due to faulty memory, as is frequently supposed, but to other factors. In an earlier paper [15] the author attempted to demonstrate that the unusual visual perception of brain-injured subjects is one factor accounting for their relative inability to reproduce visual designs. The observation holds true even when the stimulus figures are to be reproduced from copy and not from memory. The exogenous subjects' distorted reproductions of visual designs appear not to be the result of faulty memory but rather of misperception of the stimulus figures.

Using the Ellis Visual Designs Test [21] as a means of quantifying a child's drawing ability, the author noted earlier [14] that some mentally retarded children exhibited difficulty in changing their responses appropriately when shifting from one drawing to the next. There was a tendency for some of the subjects to produce the same or similar drawings in response to radically different stimulus figures, which might be called a tendency to perseverate. These perseverative responses seemed to be more common for the brain-injured than for the familial defectives. It was therefore thought possible that a further factor contributing to the defective drawing performance of the exogenous subjects might be their greater tendency toward rigidity in conceptual thinking.

[1] The writer gratefully acknowledges the cooperation of the superintendent and staff of the Ontario Hospital School, Orillia, Ontario, for making available the facilities for this study.

In accordance with Goldstein [5], Werner [20], and others, rigidity in the present study is defined as the tendency of an individual to resist change maladaptively. Perseveration is regarded as one type of rigid response that repeats itself maladaptively. Flexibility is regarded as the capacity of an individual to alter mental sets and corresponding responses appropriately as these are required by changes in the stimulus situation.

A number of investigators have found evidence that the brain-injured are characterized by rigidity. In 1928 Cameron [2] showed that brain-operated rats are less able than healthy animals to break a habit and form a new one. Maier [16] has demonstrated that operated rats tend to perseverate, that is, to repeat the same errors in successive trials. Goldstein [6, 7], in more recent studies with brain-injured and normal adults, concluded that rigidity is a general pathological trait that is present irrespective of the focus of lesion.

Rigidity is not, however, an exclusive characteristic of the brain-injured. It has been demonstrated that mentally retarded individuals whose deficiency is endogenous also show rigidity. Werner [19] refers to their rigidity as *subnormal,* in comparison with that of the brain-injured child which is termed *abnormal* rigidity. One example of a study made of subnormal rigidity is that of Lewin [12], who found that endogenous as compared with normal subjects displayed a significantly higher degree of rigidity. The investigations of Strauss and Werner [18] and Carrison and Werner [3], dealing with abnormal rigidity, have shown that exogenous as compared with endogenous defective children exhibit perseveration and a general difficulty in shifting from one task to another.

Many research workers, including Goldstein

and Scheerer [8], Weigl [19], Kasanin and Hanfmann [11], and Werner [19, 20], have devised tests in an attempt to measure the presence or absence of rigidity. After exploring the applicability of these tests on small groups of mentally defective children, the author found them generally inappropriate. The instrument finally selected for use in the present study was a modification of a Weigl-type card-sorting test first described by Berg [1], and later by Grant and Berg [9], at the University of Wisconsin. The test is an objective technique for measuring human abstraction and shift of set and is claimed by its authors to yield quantitative measures of these aspects of thinking and reasoning.

The Investigation

Problem. The purpose of this study is to investigate the hypothesis that exogenous as compared with endogenous mentally retarded children show greater rigidity in conceptual thinking.

Subjects. Two groups of mentally defective subjects, exogenous and endogenous, more fully described in a previous publication [15], were chosen from a total population of 2,000 institutionalized patients. The 15 exogenous subjects included all of the available patients who satisfied the rigid selection criteria. The 15 endogenous subjects, six boys and nine girls, were paired as closely as possible with the exogenous subjects for sex, chronological age, and intelligence. Intelligence was measured using the Revised Stanford-Binet Intelligence Scale, Form M. The two groups were composed of subjects between the ages of 10 and 20 with IQ's between 40 and 75. Subjects were selected by a member of the medical staff, a geneticist, and the author. The presence or absence of brain injury in each subject was established (*a*) by examination of case history material and (*b*) by neurological examinations of the subjects themselves.

Method. The modified Wisconsin card-sorting test used in the present study consists of a pack of 27 response cards made of white cardboard and approximately 3½ × 5 inches in size. Each card contains from one to three identical figures of a single color. Three kinds of figures—diamonds, squares, and triangles—and three colors—green, blue, and red—are used. A single card might have one green diamond, or two blue squares, or three red triangles, or any such combination. Thus, the cards may be sorted according to any one of three different principles (*a*) the color, (*b*) the form, or (*c*) the number of figures on each card. In addition to the 27 response cards which are sorted by each subject, there are three stimulus cards which are placed on the table by the examiner as guides for sorting. The three stimulus cards are always placed in the following order: the card showing one green diamond; the card showing two blue squares; and the card showing three red triangles.

When the cards are first presented to the subject for sorting, they are arranged in a standard order, but for any subsequent sortings the cards are simply shuffled. Sorting in terms of color is the principle first accepted as correct. The subject, of course, receives no instructions to this effect. He is simply told that his response is incorrect when he is not placing the response cards in front of the stimulus cards according to color. After a subject has sorted five successive cards according to the principle of color it is assumed that he has grasped the principle that the examiner accepts as correct and is using it in his sorting. At this point the examiner calls such sortings of the cards incorrect and continues to do so until the subject next learns to sort them according to either the form or the shape of the designs on the cards. The same general procedure is repeated until the three sorting principles of color, form, and number have been successfully grasped or until the pack of response cards has been sorted completely six times. The following instructions are given to each subject:

"I want you to put these cards on the table in front of one of the three cards already there (examiner demonstrates). Lay the cards down one at a time. As you can see, the cards differ in three ways: in the *color* of the spots on each card; in the *number* of spots on each card; and in the *shape* of the spots on each card. Put each card down where you think it should go and I will tell you when you are right and when you are wrong. Remember that there are three ways to sort the cards. Don't take too long but get as many right as you can."

Scoring procedure. Rigid behavior on the card-sorting test is scored according to three different criteria:

1. Perseverative responses: A perseverative response occurs when a card is sorted in terms of a principal previously accepted as correct (e.g., a card sorted according to the principle of color when it is supposed to be sorted according to the principle of form).

2. The total number of cards sorted (correct-

ly or incorrectly) before each of the three sorting principles is grasped.

3. The time required to grasp and apply each new sorting principle.

Results and Discussion

The exogenous as compared with the endogenous mentally retarded subjects showed, on the average, significantly more rigid behavior on the modified Wisconsin card-sorting test. As indicated in Table 1, rigidity was apparent not only in the subjects' total number of perseverative responses, but also in the other two measures of sorting efficiency, that is, in the total number of cards sorted and in the time needed to grasp and apply the three sorting principles.

Table 1

Group Difference in Total Rigidity Scores on the Modified Wisconsin Card-Sorting Test

Marking criteria	Endogenous subjects ($N=15$)	Exogenous subjects ($N=15$)	r†
Color			
Perseverative score	0	0	
Number of cards sorted	198	311	1.56
Time in seconds	1544	2917	1.80
Form			
Perseverative score	23	160	2.55*
Number of cards sorted	210	580	3.70**
Time in seconds	1551	3971	2.95*
Number			
Perseverative score	52	248	4.91**
Number of cards sorted	372	915	4.71**
Time in seconds	2719	5993	2.91*
Total scores			
Perseverative score	75	408	5.38**
Number of cards sorted	780	1793	6.17**
Time in seconds	5715	13035	4.31**

† The t ratios were calculated using Weaver's variation [13] of the usual formula for small matched groups.
* Significant at 5% level of confidence.
** Significant at the 1% level.

As mentioned in the section on scoring procedure, the perseverative scores represent the number of errors made when cards are sorted in terms of the principle previously accepted as correct. This interpretation of perseveration agrees with that of Werner [19], who uses the term to indicate the persistent repetition or continuation of an activity once begun. A direct comparison of the perseveration scores of the two etiological groups showed the endogenous group mean to be 5.0 (range 2—12) and the exogenous group mean to be 27.2 (range 2 — 59). In addition to the differences between the scores of the two groups being statistically significant at the one per cent level of confidence, there was an apparent cut-off score. All of the endogenous, but only three of the exogenous, subjects had a score of 12 or less. The group differences are presented graphically in Figure 1. As was expected, a definite etiological difference in favor of the endogenous mentally retarded subjects appears to obtain.

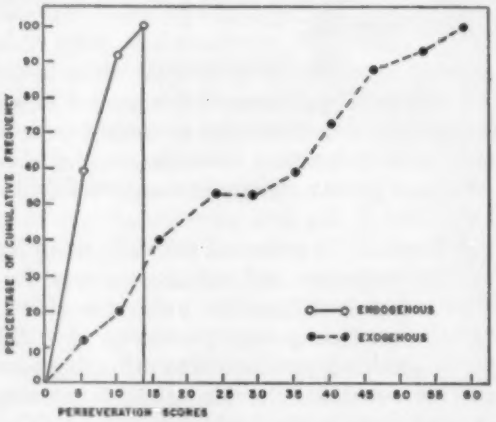

Fig. 1. Cumulative percentage distributions of perseveration scores for the endogenous and exogenous groups.

In addition to the greater number of perseverative responses made by the exogenous subjects, both the number of responses necessary and the time taken to achieve satisfactorily the three principles were significantly greater for the exogenous than for the endogenous mentally defective subjects. The longer time and greater number of responses required is considered by the author to be the result of the exogenous subjects' reduced ability to shift their "set" readily from one sorting principle to another.

In comparing the scores achieved by the two groups of subjects on the card-sorting test, it was found that both groups encountered greater difficulty as the test progressed. While it has already been shown in Table 1 that the exogenous, as compared with the endogenous, mentally retarded subjects showed more rigid behavior in the test as a whole, the groups re-

sponded similarly in that they both found sorting according to the principle of number more difficult than sorting according to the principle of form. This is probably due, in part at least, to the fact that rigidity is proportional to the number of principles by which the subjects have learned to sort the cards. The results of the present study, however, tend to support Heidbreder's [10] study performed with normal subjects, which found that of the three concepts, number was the last concept to determine a sorting. The present study is also in agreement with the investigations reported by Senden [17] on the perceptions of the congenitally blind after operation. He indicates that the concept of color persistently dominates form in the first vision of such patients. Eleven months after operation the color names that had been learned were retained, but the little that had been learned of form was forgotten.

A comparison of the rigidity scores in this investigation with the perception scores obtained in a previous study [15] denoted no significant relationship between these factors. Product-moment correlations (endogenous group + .20; exogenous group — .10) were not significant for either etiological group. This would tend to indicate that visual perception, as measured by Lissajou's figures, is relatively independent of a subject's ability to shift his set from one task to another on a card-sorting task.

Summary and Conclusions

Previous investigations [5, 14] have indicated that exogenous mentally retarded children are significantly less able than those of endogenous etiology to reproduce visual designs from copy. In a previous paper [15] the author provided evidence that the unusual visual perception of the exogenous defectives was one factor accounting for their poor drawing performance. The purpose of the present study was to investigate the hypothesis that the relative inability of the brain-injured to reproduce visual designs from copy, in comparison to familial defectives, is due to rigidity in conceptual thinking.

In the present study a modified version of the Wisconsin card-sorting test was employed to study rigidity in conceptual thinking in exogenous and endogenous mentally retarded children. The subjects from the two groups were paired on the basis of sex, chronological age, and intelligence.

It was demonstrated that the brain-injured are significantly more rigid in conceptual thinking as shown by their greater tendency to perseverate on a card-sorting task. The exogenous child is indeed less able to shift his mental set from a sorting principle such as color to another such as form or number. Such findings warrant the conclusion that the relatively poor drawing performance of exogenous defectives on such memory-for-designs tests as that found at year IX on the Binet are not due to faulty memory but to such factors as perseveration and rigidity in conceptual thinking.

Received January 14, 1954.

References

1. Berg, Esta A. A simple objective technique for measuring flexibility in thinking. *J. gen. Psychol.*, 1948, **39**, 15-22.
2. Cameron, N. Cerebral destruction in its relation to maze learning. *Psychol. Rev.*, 1928, **39**, 68-71.
3. Carrison, D., & Werner, H. Principles and methods of teaching arithmetic to mentally retarded children. *Amer. J. ment. Def.*, 1943, **47**, 281-289.
4. Cassel, R. H. Relation of design reproduction to the etiology of mental deficiency. *J. consult. Psychol.*, 1949, **13**, 421-428.
5. Gelb, A., & Goldstein, K. *Psychologische Analysen hirnpathologischer Faelle.* Leipzig: Barth, 1920.
6. Goldstein, K. The modifications of behavior consequent to cerebral lesions. *Psychiat. Quart.*, 1936, **10**, 586-610.
7. Goldstein, K. *The organism.* New York: American Book Co., 1939.
8. Goldstein, K., & Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
9. Grant, D. A., & Berg, Esta A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J. exp. Psychol.*, 1948, **38**, 404-411.
10. Heidbreder, Edna. Conceptual achievements during card sorting. *J. Psychol.*, 1949, **27**, 3-39.
11. Kasanin, J., & Hanfmann, Eugenia. An experimental study of concept formation in schizophrenia. *Amer. J. Psychiat.*, 1938, **95**, 35-48.

12. Lewin, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.
13. McCullough, M. W. Wechsler-Bellevue changes following prefrontal lobotomy. *J. clin. Psychol.*, 1950, **6**, 271-274.
14. McMurray, J. G. A study of factors contributing to poor drawing performance in exogenous as compared with endogenous feebleminded children. Unpublished master's dissertation, Univer. of Western Ontario, 1952.
15. McMurray, J. G. Visual perception in exogenous and endogenous mentally retarded children. *Amer. J. ment. Def.*, 1954, **58**, 659-663.
16. Maier, N. R. F. The effect of cerebral destruction on reasoning and learning. *J. comp. Neurol.*, 1932, **54**, 45-75.
17. Senden, M. *Raum- und Gestaltauffassung bei operierten Blindgeborenen vor und nach der Operation.* Leipzig: Barth, 1932.
18. Strauss, A. A., & Werner, H. Experimental analysis of the clinical symptom "perseveration" in mentally retarded children. *Amer. J. ment. Def.*, 1942, **47**, 185-188.
19. Werner, H. Abnormal and subnormal rigidity. *J. abnorm. soc. Psychol.*, 1946, **41**, 15-24.
20. Werner, H. The concept of rigidity: a critical evaluation. *Psychol. Rev.*, 1946, **53**, 43-52.
21. Wood, Louise, & Schulman, Edythe. The Ellis visual designs test. *J. educ. Psychol.*, 1940, **31**, 591-602.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Test Performance of Preschool Children With and Without Organic Brain Pathology[1]

Miriam Stearns Haines
The Public Schools, Montclair, New Jersey

The diagnosing of brain injury is of major importance. Until recently the responsibility for this diagnosis rested solely with the medical profession. Today it is recognized that psychological findings may contribute to the final diagnosis along with the neurological examination and such laboratory tests as the electroencephalogram. Although a correct diagnosis is important for mature individuals, it is of even greater consequence for young children because of the impact of brain injury on the developing organism [4]. Brain-injured adults retain much of what they have learned even though the acquiring of new skills is impaired. The situation is more serious, however, for children who have sustained cerebral insult before many tasks are normally learned. If a correct diagnosis is not made, they may not receive the specialized training which they need in order to learn normally expected skills. Methods have been especially designed to teach them such necessary functions as reading and arithmetic fundamentals [16]. If a diagnosis of primary amentia is made rather than of secondary amentia due to brain injury, these children are frequently placed in institutions for the endogenous mental defective and barred from an opportunity for even rudimentary academic learning [5].

Two assumptions are made when a psychological test is expected to differentiate between young children with and without brain injury: (*a*) that children's abilities are sufficiently developed and specialized at an early age for differential patterns to emerge, and (*b*) that the test is sensitive enough to tap varying degrees of ability [2]. The fact that test items in both verbal and performance areas are administered to children and infants as young as four weeks on the Gesell Developmental Schedules and at two months on the Cattell Infant Tests lends weight to the concept of differential abilities in the young child [1, 9].

On the basis of previous research with older children there is reason to suspect that young brain-injured children have difficulty in conceptual thinking, in visual-motor and perceptual judgment, and there is also the possibility of more general mental retardation [7, 13, 14, 17, 21]. Whether or not intellectual impairment follows a brain injury depends on the area of the brain affected. If the lesion is in the cerebral cortex, intellectual impairment or "deficit" is a definite possibility although even here not an inevitable consequence [18]. Literally, any case of feeblemindedness might be considered as possibly due to birth lesion when one considers the whole process of fetal development and the difficulties of childbirth.

Method

The instrument chosen in this investigation to test varying degrees of ability was the Merrill-Palmer Scale of Mental Tests [19]. It was selected primarily because it consists of a battery of verbal and performance tests that permits scoring for different degrees of ability through the preschool years. In addition, it has strong appeal for young children and very few resist "playing the games." Motivation is an important factor since cooperation is frequently difficult to obtain in brain-injured children on tests without intrinsic appeal. In fact, electroencephalograms are often reported to be unsatisfactory in these children because of lack of cooperation.

It is hypothesized that young children with brain injury will show decreased efficiency on

[1]Acknowledgment is made to the National Institutes of Health of the U. S. Public Health Service for a fellowship grant to the author to conduct this study and to Miss Louise R. Hewson, Director, Psychology Department, Neurological Institute, New York City to collect the data.

tests involving conceptual thinking, visual-motor skill, and perceptual judgment. Specifically, it is expected that the brain-injured group should show significantly lower means on the selected Merrill-Palmer tests that purport to measure these abilities than the control group. A brain-injured child in this study is defined as one who has suffered a brain lesion due to infection or injury occurring before, during, or after birth. Conceptual thinking is conceived as the ability to form a concept, to classify according to a common factor. Visual-motor skill is dependent upon eye-hand coordination, finger dexterity, speed of reaction time, spatial evaluation, and visual memory. Perceptual judgment implies the discriminative observation of differences, relationships, and organizations.

The experiment was designed for a brain-injured group of 100 children and two non-brain-injured groups consisting of 100 behavior-problem and 100 foster-home children to whom the Merrill-Palmer and the Revised Stanford-Binet, Form L, were administered. The selection of behavior-problem children for comparison with the brain-injured group was made partly because of their availability in the clinic as a non-brain-injured population but mainly because differentiation between children with brain injury and those with behavior problems is an especially important issue which the neuropsychiatrist refers to the psychologist [3, 10, 11, 12]. Comparison could have been made with so-called "normal" children in nursery school, kindergarten, and primary grades, but the foster-home group was selected instead because from the standpoint of this study there would have been several disadvantages with a normal group. First, there is screening of children for admission to these educational groups, and those with low intelligence quotients would probably have been eliminated, especially where there were physical stigmata. Second, the testing could not have been done by the same examiners at the same place, whereas identical conditions prevailed for the three groups that were selected.

The only criteria for selection of subjects besides their diagnosis were chronological age from 3 years, 0 months to 7 years, 10 months; Intelligence Quotients of at least 45 on the Revised Stanford-Binet, Form L; and basal age at least 30 months on the Merrill-Palmer [20]. These limitations stem from the selection of Merrill-Palmer tests and a choice of a predominantly preschool population. The mean Stanford-Binet IQ for the brain-injured and foster-home children was found to be 93 and for the behavior-problem group 89. Although no attempt had been made to match the three groups for sex, the brain-injured were found to have a one-to-one ratio, while both the non-brain-injured groups had a higher proportion of boys. None of the children had visual or auditory impairments, nor was there a difference in the socioeconomic backgrounds.

The etiologies in the exogenous group were: birth injury, 5; meningitis or encephalitis, 15; hypoplasia of the brain, 23; tumor of the brain, 10; grand or petit mal epilepsy, 16; cortical seizures, 11; hydrocephalus, 9; brain-injury etiology unknown, 11. The electroencephalograms and pneumoencephalograms were abnormal. The behavior-problem and foster-home children are known as non-brain-injured groups because their developmental histories and neurological examinations gave no evidence of a lesion of the central nervous system.

The tests selected were the Action Agent, the Buttons, the Copying a Circle, Cross, and Star, the Seguin Formboard, the Pyramids, the Pink Tower, the Mare and Foal, and the Picture Puzzles because they tap a variety of abilities as well as provide for varying degrees of proficiency among children from 30 to 77 months. The tests usually administered below 30 months were omitted because they are based on the all-or-none principle of scoring; for example, the child can either cut with a pair of scissors or he cannot; he can either identify himself in a mirror or he cannot.

An analysis of the abilities in the Merrill-Palmer was made as a preliminary step so that the abilities involved in each test were defined prior to the analysis of the relevant data. Three competent psychologists who had clinical experience with the Merrill-Palmer scale and had taught and demonstrated its technique to graduate students and externs were chosen to check on the writer's analysis of abilities. The judges were sent the writer's analysis of the abilities involved in the Merrill-Palmer tests together with a list of definitions for the abilities. The combined efforts resulted in a grouping of the tests in terms of the abilities in each test as they related to the hypothesis under investigation. For example, under conceptual thinking was listed the test which involves this ability

primarily; under visual-motor skill those tests which embody spatial evaluation, speed of reaction time, and visual memory; and under perceptual judgment tests which are predominantly dependent upon this ability.

Results

The Merrill-Palmer mean scores for each of the three diagnostic groups ranged from 3.0 to 4.8, where there is a possible range from 0 to 7.0. When the *t* test of the significance of the differences between the means of the three groups, for each subtest, was computed, it was found that no test was significantly different at the .05 level in all instances. When the *t* test was applied to all differences between subtest means within each group, no one test was found to reach significance at the .05 level in all seven instances. Thus, the null hypothesis must be accepted in terms of the criterion, and the differences among the three groups and between tests in any one of the groups must, therefore, be attributed to chance. Table 1 shows the *t* tests between the means of the eight Merrill-Palmer tests for each pair of diagnostic groups.

Discussion

The results of this investigation do not demonstrate that young children with brain injury show decreased efficiency on tests presumably measuring conceptual thinking, visual-motor skill, and perceptual judgment. This may be because abilities at this young age level are general rather than specific [8]. Other studies which report significant differences dealt with older children and adults. It is also possible that the Merrill-Palmer as a measuring device is not a sensitive enough instrument to differentiate degrees of conceptual thinking, spatial evaluation, and perceptual judgment compared to those tests which have been specifically designed for differential diagnosis. The fact that little relationship was found between Stanford-Binet IQ's and Merrill-Palmer test scores tends to support this interpretation. The present research is in line with the recent attempts to explore the contributions of intelligence tests to neuropsychiatric diagnosis. Since these tests were designed to measure global intelligence and are based on theory directed toward that end, it should be no disparagement of the Merrill-Palmer that it does not perform a function foreign to its basic premise [15].

Summary

This study investigated the ability of selected subtests of the Merrill-Palmer scale to differentiate between brain-injured and non-brain-injured children, three to seven years of age. The experimental group of 100 brain-injured children was compared to two control groups of 100 behavior-problem children and 100 foster-home children. Analyses were made by *t* tests of the differences of mean subtest scores within each group. Although a few scattered *t* tests showed statistical significance, no test pattern emerged that validly differentiated the brain-injured from the other groups. The null hypothesis must be accepted. It is concluded that subtests of the Merrill-Palmer scale are not useful for the differential diagnosis of brain-injured children from behavior-

Table 1

t Tests between Means of Eight Merrill-Palmer Tests for Each Pair of Diagnostic Groups

Groups	Tests							
	Seguin	Picture Puzzles	Buttons	Action Agent	Pink Tower	Mare & Foal	Pyramids	Circle, Cross, Star
Brain-injured vs. behavior problem	1.27	1.66	.52	1.76	1.84	.90	1.74	.92
Behavior problem vs. foster home	2.21*	2.70**	.23	3.02**	1.49	.83	1.49	1.60
Brain-injured vs. foster home	.88	1.01	.78	5.06**	.30	.06	.22	.58

* Significant at .05 level of confidence (1.96).
** Significant at .01 level of confidence (2.576).

problem or foster-home children.

Received January 28, 1954.

References

1. Cattell, Psyche. *The measurement of intelligence of infants and young children.* New York: Psychological Corp., 1947.
2. Clark, M. P. Changes in primary mental abilities with age. *Arch. Psychol.,* 1944, No. 291.
3. Cobb, S. Personality as affected by lesions of the brain. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 550-581.
4. Doll, E. A. Behavioral consequences of cerebral birth lesions. *Amer. J. Obst. Gynecol.,* 1936, **31,** 866.
5. Fay, Temple, & Doll, E. A. Organic impairment simulating mental deficiency. *Amer. J. Orthopsychiat.,* 1949, **19,** 112-119.
6. Fisher, R. A. *Statistical methods for research workers.* (6th Ed.) London: Oliver & Boyd, 1936.
7. Frazeur, Helen A., & Hoakley, Z. Pauline. Significance of psychological test results of exogenous and endogenous children. *Amer J. ment. Def.,* 1947, **51,** 384-388.
8. Garrett, H. A developmental theory of intelligence. *Amer. Psychologist,* 1946, **1,** 372-378.
9. Gesell, A., & Amatruda, Catherine S. *Developmental diagnosis.* (2nd. Ed.) New York: Hoeber, 1948.
10. Halstead, W. C. *Brain and intelligence.* Chicago: Univer. of Chicago Press, 1947.
11. Kasanin, J. Personality changes in children following cerebral trauma. *J. nerv. ment. Dis.,* 1929, **69,** 385-406.
12. Lashley, K. S. *Brain mechanisms and intelligence.* Chicago: Univer. of Chicago Press, 1929.
13. Sarason, S. B. *Psychological problems in mental deficiency.* New York: Harper, 1949.
14. Sarason, S. B., & Sarason, Esther K. The discriminatory value of a test pattern with cerebral palsied defective children. *J. clin. Psychol.,* 1947, **3,** 141-147.
15. Shaffer, L. F. Theoretical bases for psychometric tests. In P. H. Hoch & J. Zubin (Eds.), *Relation of psychological tests to psychiatry.* New York: Grune & Stratton, 1950. Pp. 26-38.
16. Strauss, A. A., & Lehtinen, Laura E. *Psychopathology and education of the brain-injured child.* New York: Grune & Stratton, 1947.
17. Strauss, A. A., & Werner, H. Disorders of conceptual thinking in the brain-injured child. *J. nerv. ment. Dis.,* 1942, **96,** 153-172.
18. Strecker, E. A., & Ebaugh, F. G. Neuropsychiatric sequelae of cerebral trauma in children. *Arch. Neurol. Psychiat.,* 1924, **12,** 443-453.
19. Stutsman, Rachel. *Mental measurement of preschool children.* Yonkers, N. Y.: World Book Co., 1931.
20. Terman, L. M., & Merrill, Maud A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.
21. Werner, H., & Strauss, A. A. Impairment in thought processes of brain-injured children. *Amer. J. ment. Def.,* 1943, **47,** 291-295.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Sex Differences on the Revised Beta Examination

George Lavos
Michigan School for the Deaf

The Revised Beta Examination is a nonlanguage test measuring "general intellectual ability" for those 16 years of age and older [1, p. 2]. No reports of its use with the deaf appear in the literature although it is adaptable for use with the deaf in administration, format, and nonlanguage response.

The Problem

The Revised Beta Examination is provided with norms based on white male adults. These men were selected so that they would represent proportionately in the normative population the educational and socioeconomic status of the adult white males in the 1940 U.S. census. Although the standardization group is limited to men, the Manual states that "the Examination has been used for testing women as well as men for many years and no report has been found indicating that different norms were required for the different sexes" [1, p. 6].

The assumption that the norms based on men are appropriate for women should be tested before use of the examination is recommended for deaf women and older girls. The problem is, then, when used with deaf youths is there a sex difference in the intelligence quotients based on the Revised Beta Examination that cannot be attributed to chance sampling in the population?

Method

The records of 188 deaf pupils, between the ages of 16 and 19 at the time of testing, were available. The average age at the time of administering the test was 17-3. The pupils were accustomed to testing since the Beta test was the third group test of intelligence used in the school from which the data were obtained. There were 79 girls and 109 boys in the group. All were regularly enrolled pupils in the school and had severe hearing losses.

The intelligence quotients for the total group ranged from 62 to 128. These quotients were calculated according to the Lindner-Gurvitz procedure recommended for use with the Beta test.

In order to find out whether there is a sex difference on the Beta, it is necessary to ascertain whether the performance of the two sexes is the same on similar tests. Since the Beta was the third of a series of tests of nonlanguage intelligence used with the subjects of the study, results of the first two tests could be used to assess the possibility that a sex difference already exists in the group. The critical ratio of the difference between the means of the quotients for boys and girls on the first of the two tests given earlier, the Chicago Non-Verbal Examination, was 1.44. The critical ratio for the mean quotients from the second test, the Pintner Non-Language Test of Mental Ability, Intermediate Series, was .005. Both ratios indicated that no sex differences exist on the Chicago and the Pintner tests that cannot be attributed to chance factors.

Results

The mean Beta intelligence quotient of the girls (94.4) was lower than that of the boys (103.0). The critical ratio of the difference in means was 4.34. The probability is less than 1 in 1,000 that a critical ratio this large can arise by chance. There is a reliable difference in the means of deaf boys and deaf girls on the test; consequently, the two sexes cannot be considered to be samples from a common population with respect to the capacities measured by the test.

In the light of this finding it is desirable to see on which subtests of the Beta deaf boys and girls differ significantly in performance. By using the weighted scores for each subtest that are provided in the Lindner-Gurvitz standardization, it is possible to compare directly any person within the age range from 16 to 19, 20

to 24, and so forth, without considering growth. Thus the boys and the girls, each group ranging in age from 16 to 19, can be directly compared without regard to variations in age within that range.

The differences in means of the weighted scores for the boys and the girls on each of the six subtests of the Beta were tested for significance. Three mean differences were statistically significant, and these were in favor of the boys. The three subtests in which significant differences appeared were "error recognition," "paper form board," and "picture completion."

The subject matter of the drawings in the items for "error recognition" is largely masculine. The "paper form board" subtest is similar to those appearing in mechanical aptitude tests. In "picture completion" there is no preponderance of masculine or feminine orientation.

Summary

In order to test the assumption that the norms for the Revised Beta Examination can be used with deaf women, the norms being derived from hearing males only, the test results of 79 deaf girls and 109 deaf boys between 16 and 19 years of age were analyzed. It was shown that the performance of these subjects was not indicative of a sex difference in intelligence quotient on the Chicago Non-Verbal Examination and the Pintner Non-Language General Ability Test, which had previously been given to the subjects. On the Beta, however, a significant sex difference in mean intelligence quotient existed in favor of the boys. Subtest analysis located the difference in three subtests, two of which are masculine in orientation. It is concluded that the Beta test cannot be used with deaf girls of 16 to 19 years of age without appropriate norms.

Received February 9, 1954.

Reference

1. Lindner, R. M., & Gurvitz, M. *Revised Beta Examination manual.* New York: Psychological Corp., 1946.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

A Validation of a Bender-Gestalt Scoring System [1]

Murray Lonstein
VA Hospital, Lebanon, Pennsylvania

The Bender-Gestalt test (B-G) has become a part of test batteries in many hospitals and clinics throughout the country. Its need for quantification and validation has been recognized, and, accordingly, scoring systems have been devised by several investigators [1, 4, 5]. One such scoring system, constructed by Pascal and Suttell [4], appears worthy of further study in view of the fact that its authors have claimed that the method discriminates normal subjects from neurotics and psychotics. Furthermore, Robinson [6] found that this system differentiates the B-G performances of paretics and schizophrenics, and Cornutt [2] supported the claims of Pascal and Suttell in regard to scoring and collection of their normative data.

The present study was designed to validate further the Pascal-Suttell scoring system and to establish its practical usefulness in a hospital setting. It was held that even if the system is not sensitive enough to distinguish neurotic from normal subjects, as Gobetz [3] claimed, its efficacy in separating psychotic from nonpsychotic patients would merit its continued use.

Procedure

A total of 96 B-G records of psychiatric patients in a veterans hospital was scored by the author according to the directions given by Pascal and Suttell. Briefly, this technique involves an analysis by the psychologist of Designs 1-8, which are scored according to criteria given in their manual. The number of points for each specified deviation increases with its severity. The sum of these points is the raw score, which is converted into a standard measure or z score.

After the records were scored, the psychiatric diagnoses of the patients were obtained and the records were placed into two groups, "psychotic" and "nonpsychotic." In the psychotic group there were 41 schizophrenics, 4 psychotics with organic brain disease, and 2 manic-depressives. In the nonpsychotic group there were 40 anxiety reactions, 6 passive-aggressive reactions, 1 obsessive-compulsive reaction, 1 dissociative reaction, and 1 alcoholic.

Since the Pascal-Suttell z scores are based on norms that exclude subjects whose education is below the ninth grade, 23 patients who did not meet this standard were not used in the z-score comparison. They were included, however, in the raw score analysis.

Results

As indicated in Table 1, the means of the standard score distributions of the two groups differed significantly; likewise, the means of the larger raw score distributions also differed significantly. These findings indicate that the Pascal-Suttell scoring system meets the statistical criterion in differentiating psychotic from nonpsychotic patients, using either raw or standard scores.

Table 1

Statistical Comparisons of B-G Standard and Raw Scores of Psychotic and Nonpsychotic Groups

Statistic	Standard scores		Raw scores	
	Psychotic	Non-psychotic	Psychotic	Non-psychotic
N	36	37	47	49
M	72.19	56.05	35.30	23.96
SD	22.48	14.12	22.56	14.53
CR	3.67 ($p < .001$)		2.92 ($p < .01$)	

Although the null hypothesis may be rejected with a high degree of confidence, it was deemed more appropriate to evaluate this technique as a practical method by determining from the data the percentage of true and false positives. In order to satisfy this important criterion, it was necessary to find appropriate cutoff points to separate the two groups. Pascal and Suttell report that z scores between 50 and

[1] From the VA Hospital, Lebanon, Pennsylvania.

72 are suspect for psychiatric illness. A z score of 60, one standard deviation above their normal group mean of 50, is suggested by them as a reasonable cutoff score in differentiating between patients and nonpatients.

Inspection of the data reveals that of the 36 psychotic records on which z scores were computed, 24 patients had scores greater than the suggested cutoff score of 60, and 12 had scores below 60. As indicated in Table 2, 66.66 per cent of the psychotics were correctly identified. Of the 37 nonpsychotic patients, 13 had scores above 60, whereas 24 had scores below 60; that is, 64.86 per cent were correctly identified. In the total sample of 73 patients, the percentage of true positives was 65.76 and the percentage of false positives was 34.24.

Table 2

Percentage of True and False Positives with Cutoff B-G Scores of 60 and 65

Percentages	Cutoff score at 60	Cutoff score at 65
% psychotics correctly identified	66.66	63.89
% nonpsychotics correctly identified	64.86	73.38
Total % false positives	34.24	28.77
Total % true positives	65.76	71.23

Since all the subjects in this study were hospitalized and therefore maladjusted to some degree, it seemed reasonable to select a somewhat higher cutoff score; accordingly, a cutoff score of 65 emerged as a more effective score for reducing the total percentage of false positives. Only 8 of the 37 nonpsychotic records had z scores over 65, and 13 of the 36 psychotics had scores below 65; that is, with the higher cutoff score, 1 additional psychotic was misidentified, but 5 more nonpsychotics were correctly identified. The percentages in Table 2 reflect this trend and show the total percentage of true positives as 71.23, and the total percentage of false positives as 28.77. In other words, the technique correctly separated psychotics from nonpsychotics in about 7 out of 10 cases when a cutoff score of 65 was used.

Since all of the B-G performances were timed when the test was administered, it was believed that knowledge of possible differences between the two groups on the time variable would be helpful diagnostically. Accordingly, the time measures (seconds taken to copy all of the designs) were compared on a group basis. The psychotic group's mean time measure was 445.03 seconds, with a sigma of 290.25 seconds; the nonpsychotic group averaged 331.05 seconds, with a sigma of 112.82 seconds. The critical ratio derived in this comparison was 2.29, significant beyond the .05 level of confidence—that is, psychotic patients worked more slowly on the test than did nonpsychotic patients.

Discussion

Although the Pascal-Suttell scoring system meets a high statistical level of confidence in differentiating between hospitalized psychotic and nonpsychotic patients, it is by no means infallible in the practical, clinical setting. Nevertheless, in practical work it can discriminate between these two groups on a better than chance basis. The variables selected for scoring are therefore of diagnostic importance and perhaps should be studied by those who use this test even though they may not formally score each B-G record. However, since the scoring procedure can be learned easily and a record can be scored quickly, the technique should be of practical usefulness in clinical settings.

As Pascal and Suttell point out, an individual z score derived by use of their system must be interpreted with caution. Although a high z score suggests psychiatric illness, it does not always mean that the subject is psychotic. Some factors incidental to the testing situation, such as fatigue, may contribute to high scores. Like all such test measures, the z score must be viewed along with other test scores and clinical data. Furthermore, the fact that the criterion used in this study is psychiatric diagnosis imposes some limitations on the reliability of these results.

Summary

A validating study was carried out on the Pascal-Suttell scoring system of the Bender-Gestalt test. The records of 96 psychiatric patients were scored, and it was found that this technique differentiated psychotic from nonpsychotic records by meeting statistical criteria and, in a limited way, practical criteria. An in-

cidental finding is that psychotic patients took more time in reproducing the designs than did nonpsychotic patients.

Received December 16, 1953.

References

1. Billingslea, F. Y. The Bender-Gestalt: an objective scoring method and validating data. *J. clin. Psychol.*, 1948, **4**, 1-27.
2. Cornutt, R. H. The use of the Bender-Gestalt with an alcoholic and non-alcoholic population. *J. clin. Psychol.*, 1953, **19**, 287-290.
3. Gobetz, W. A quantification, standardization, and validation of the Bender-Gestalt test on normal and neurotic adults. *Psychol. Monogr.*, 1953, **67**, No. 6 (Whole No. 356).
4. Pascal, G. R., & Suttell, B. J. *The Bender-Gestalt test.* New York: Grune & Stratton, 1951.
5. Peek, R. M., & Quast, W. *A scoring system of the Bender-Gestalt test.* Hastings, Minn. (Box 292) and Minneapolis, Minn. (2810 42nd St.): Authors, 1951.
6. Robinson, N. M. Bender-Gestalt performances of schizophrenics and paretics. *J. clin. Psychol.*, 1953, **19**, 291-293.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Gellerman's Complex Crossover Design[1]

Julian C. Stanley[2]
Department of Education, University of Wisconsin

In discussing the analysis of his Form I and II Arthur scores, Gellerman [1, p. 129] states: "The *F*-tests were designed after the manner described by Grant . . ." [2]. Yet on page 127 he indicates that assignment to the two counterbalanced order-of-testing groups (I-II, II-I) was not random: "Subjects were selected by the matched-pair method to insure minimal group differences. The factors controlled in selection were sex, chronological age, mental age, previous experience with the Arthur scales, and etiological classification."

Instead of partitioning his total sum of squares into five parts (orders, individuals within orders, sessions, forms, and individuals × sessions within orders), Gellerman could have performed the following more extensive analysis: sexes, pairs within sexes, individuals within pairs, sessions, sessions × sexes, sessions × pairs within sexes, forms, forms × sexes, and forms × pairs within sexes.

Gellerman's data are no longer available, so it is impossible to determine definitely whether or not his conclusions would be altered by the complex breakdown. To illustrate the two designs and their analyses in considerable detail, the present writer divided 28 students randomly into two groups of 14 each. One group received the two forms of a locally prepared final examination in A-B order; the other group took Form B first. Form A proved significantly harder than Form B.

Then Gellerman's design was simulated by *artificially* assigning "sexes" and by matching like-sexed students across orders. The "error" mean square for form differences dropped from 11.4 to 4.9. These were *not* Gellerman's scores, but it appears reasonable to surmise that his MA difference would also have been found more significant, rather than less so, if he had employed the longer analysis. No other comparisons were basic to his study.

Gellerman need not have matched subjects, or else he might well have taken matching explicitly into account. The basic crossover design is much easier to plan and analyze. Therefore, despite greater possible efficiency of the nine-part analysis, the simpler one seems preferable unless added comparisons themselves are of interest.

Brief Report
Received March 15, 1954.

[1]An extended report of this study may be obtained without charge from Julian Stanley, Education Building, University of Wisconsin, Madison, or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4306 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2]The writer is indebted to V. N. Amble, E. E. Cureton, A. L. Edwards, P. O. Johnson, O. Kempthorne, E. F. Lindquist, and W. J. Moonan for helpful comments concerning an earlier draft of the full report.

References

1. Gellerman, Saul W. Forms I and II of the Arthur Performance Scales with mental defectives. *J. consult. Psychol.*, 1952, **16**, 127-131.
2. Grant, David A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427-442.

Journal of Consulting Psychology
Vol. 18, No. 5, 1954

Mental Age Equivalents for the Revised Stanford-Binet Vocabulary Test[1]

Edward E. Cureton
University of Tennessee

Neither Terman and Merrill [3] nor McNemar [2] give a table of mental age equivalents for the vocabulary test of the 1937 Revised Stanford-Binet Tests of Intelligence. McNemar presents a table giving the mean and standard deviation of the vocabulary scores for each age group of the standardization sample from 7 to 18 [2, p. 141, Table 44], but says: "We do not deem it necessary or desirable to present a table of mental-age equivalents . . ." [2, p. 140]. This judgment is apparently based on the fact that the vocabulary test, which consists of only 45 items, is not highly reliable.

Both McNemar, and Terman and Merrill note that the vocabulary test has higher correlations with total score than any other. Examiners frequently use it as a brief verbal scale, moreover, so it seems worth while to make available to them the best mental age equivalents that can be derived from the published standardization data.

Derivation of Equivalents

The first column of Table 1 lists the age groups in which the vocabulary test is used in MA determination. The number of cases in each group from 6 to 14 is a little over 200, and in the groups from 15 to 18 it is a little over 100. All children in the standardization groups were within one month, plus or minus, of a birthday, so for practical purposes each age group may be taken to consist of children all of whom are exactly that age.

Column 2 of Table 1, taken from McNemar's Table 44, gives the mean vocabulary scores for age groups 7 to 18. Why he did not give the average for age 6 is not clear, as vocabulary is one of the 6-year tests of Form L. Since these mean scores vary irregularly from their general trend because of sampling errors, is was necessary to smooth them. This was done in two steps. First, a five-point weighted moving average, which preserves linear, parabolic, and cubic trends over each set of five points but removes irregularities of higher order, was applied [1, p. 374, Formula 29.7, $a_0 = 5$]. Since minor irregularities may generate a spurious cubic trend over as many as five points, this formula tends slightly to under smooth, and it leaves unsmoothed the two points at each end of the series. A seven-point formula would have given better smoothing in the middle range, but it would have left three points at each end (half the points in the series) unsmoothed. In the second step, the first

Table 1

Raw and Smoothed Mean Vocabulary Scores and Mean Composite Mental Ages of Age Groups 6 to 18 Inclusive, Revised Stanford-Binet Tests of Intelligence

Age group	Mean vocabulary score		Mean mental age		
	Raw	Smoothed	Raw	Smoothed	Smoothed
6		5.8*	6.09	6.09	6–1
7	7.0	7.0	7.19	7.19	7–2
8	8.3	8.3	8.23	8.29	8–3
9	10.0	9.8	9.43	9.39	9–5
10	11.4	11.6	10.47	10.49	10–6
11	13.7	13.5	11.49	11.54	11–6
12	15.4	15.5	12.39	12.51	12–6
13	17.4	17.1	13.50	13.33	13–4
14	18.2	18.5	13.80	14.01	14–0
15	20.0	19.8	14.73	14.66	14–8
16	21.2	21.0	15.24	15.29	15–3
17	22.0	21.9	15.76	15.77	15–9
18	22.5	22.5	16.09	16.09	16–1

* Extrapolated.

[1]Revision of a paper presented to the Tennessee Psychological Association at Nashville, Tennessee, January 23, 1954.

differences of the five-point-smoothed series, including the two unsmoothed points at each end, were set down, and a few additional changes of one digit after the decimal point were made in such a manner as to regularize these differences. From the figures at the top of the smoothed series and of the difference table, one additional value was added to the series by extrapolation at age 6. Column 3 of Table 1 gives the smoothed series.

Terman and Merrill give the Form L and Form M mean IQ's for all age groups of the standardization sample [3, p. 35, Table 5]. The two mean IQ's for each group were averaged and then multiplied by the age, or by the equivalent chronological age divisor in the case of age groups 14 to 18 [3, p. 31, Table 4], to obtain the mean L–M composite mental age. These means are given in Column 4 of Table 1. The same smoothing procedure was applied to them also. In the second stage, no change as great as one digit in the first decimal was made. The smoothed values, in years and hundredths, are given in Column 5 of Table 1. These values were then converted to years and months as shown in Column 6.

The entries in Columns 3 and 6 of Table 1 were then plotted on a large sheet of 1/10-inch, cross-section paper, using 1/2 inch per unit of the vocabulary score and 1/10 inch per month of mental age. It was noted that the trend between the points for ages 10 to 18 was almost exactly linear.

McNemar reported that two subjects, one at age 17 and one at age 18, obtained the highest vocabulary score, 38 [2, p. 140]. This performance was judged by the writer to be approximately equivalent in rarity to IQ 160, or adult mental age 24-0. One point was therefore added to the chart at score 38 and MA 24-0. A straight line drawn through this point and the point for age group 10 passed exactly through the point for age group 18, and no one of the points for age groups 11 to 17 deviated from it by as much as 1/16 inch on the scale used. Hence an upward extrapolation in a straight line from the 10-year to 18-year trend would give exactly the same result as the interpolation to the point at score 38 = MA 24-0. This exact agreement is in considerable part coincidental, of course, but it gives us some further confidence in the accuracy of the mental age equivalents of scores higher than that of the 18-year age group average.

The line was continued downward in a smooth curve through the averages for age groups 6 to 9, and extended one score point further downward and two score points further upward.

The mental age corresponding to each score from 5 to 40 was then read from the curve. Accuracy of reading was checked by forming a difference table of the mental age equivalents, and a few changes of one month were made in such a manner as to regularize these differences. The resulting mental age equivalents are given in Table 2.

Table 2

Mental Age Equivalents of Raw Scores on the 1937 Revised Stanford-Binet Vocabulary Test

Score	MA	Score	MA*
5	5–4	23	16–4
6	6–3	24	16–10
7	7–2	25	17–4
8	8–0	26	17–10
9	8–10	27	18–4
10	9–7	28	18–11
11	10–2	29	19–5
12	10–9	30	19–11
13	11–3	31	20–5
14	11–9	32	20–11
15	12–3	33	21–5
16	12–9	34	21–11
17	13–3	35	22–5
18	13–9	36	23–0
19	14–3	37	23–6
20	14–9	38	24–0
21	15–3	39	24–6
22	15–10	40	25–0

* Values in this column are interpolated to a single point: score 38 (highest made in the standardization group; N = 2) taken as equivalent to MA = 24 – 0 (IQ = 160). This interpolation corresponds perfectly to an extrapolation upward of the curve from scores 12 to 22.

Reliability

Though the mental age equivalents are given in years and months to conform to current practice, the reliability of the test does not warrant this degree of fineness. No direct data on reliability are available, but rough estimates may be made. Terman and Merrill give the median L–M correlation for the ages above 6 as .93 [3, p. 47]. They give the average correlation between vocabulary and the L–M composite as .81 [3, p. 302]. If we

raise the .93 by the Spearman-Brown formula, we obtain .964 as the estimated reliability of the L–M composite. Dividing .81 by the square root of .964 gives .825 as an estimate of the correlation between vocabulary and "true" mental age. The lower limit of the reliability of the vocabulary test is the square of .825 or .68. Its reliability could be this low only if it measured exactly the same mental function as the L–M composite, which is clearly not the case. We shall take .75 more or less arbitrarily as an estimate of its reliability in a single age group.

The standard error of measurment is the standard deviation of the scores times the square root of one minus the reliability. Thus $\sqrt{1 - .75} = .5$, so the standard error of measurement is approximately half the standard deviation of the scores. McNemar's Table 44 [2, p. 141] gives the standard deviation of the vocabulary scores for each age group from 7 to 18. Multiplying half of each of these values by the difference between adjacent mental ages in Table 2, at the point corresponding to the entry in Column 6 of Table 1 for the given age group, we obtain the standard errors of measurement in months of mental age. Converting these values to fractions of a year and dividing by the age of each group, or by the equivalent chronological age for age groups 14 and over, we obtain the standard error of measurement of the vocabulary IQ for each age group. The values obtained range from 8.67 to 11.90, with an average of 9.72. We conclude, therefore, that the standard error of measurement of the vocabulary IQ is roughly ten points. This means that in interpreting such IQ's we can place some confidence in the first figure, or the first two figures if the IQ is 100 or over, but none in the last figure. The 95 per cent confidence limits for a given vocabulary IQ are approximately plus or minus 20 points.

Summary

Table 2 gives mental age equivalents of scores on the vocabulary test of the 1937 Revised Stanford-Binet Tests of Intelligence. Any inaccuracies in the table, whether due to the method of its derivation or to the limitations of the original standardization data, are small in comparison to the unreliability of the test.

Despite its limited reliability, which is inevitable in a 45-item test covering the whole range from age 6 to superior adult, this vocabulary test is useful as a short verbal scale.

Vocabulary IQ's may be computed from these mental ages by dividing them by CA up to age 13, and thereafter by the CA equivalents used with the 1937 Revised Stanford-Binet Tests of Intelligence. For the interpretation of any such IQ, its last digit has no significance; the standard error of measurement is about 10 points.

Received January 12, 1954.

References

1. Kendall, M. G. *The advanced theory of statistics.* Vol. 2. London: Griffin, 1948.
2. McNemar, Q. *The revision of the Stanford-Binet Scale.* Boston: Houghton Mifflin, 1942.
3. Terman, L. M., & Merrill, Maud A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.

NEW BOOKS AND TESTS

Books

Arnold, Magda B., & Gasson, John A. *The human person.* New York: Ronald, 1954. Pp. x + 593. $5.75.

Ten Catholic psychologists met in a workshop "to formulate an integrated theory of personality based on a Christian conception of human nature." This volume reports the planned individual papers and extra chapters on the discussion provoked by them. The material is organized into five parts: theory building, personality structure, personality integration, psychotherapy and self integration, and self integration through religion. The chapter devoted to Frankl's existential analysis is an addition to the American literature on psychotherapy. Logotherapy or emphasis on the client's ideals, aspirations, and deliberate choices deserves detailed systematic treatment. Too often therapeutic theory and practice have neglected a strong emphasis on the client's life as an expression of unique goals with appropriate unique means to accomplish them. The editors present this book as a primary text in personality. It is rather a book on personality theory fairly wide in scope but with relatively little of the empirical findings woven into it. The two chapters on orthodox Christianity and 40 pages devoted to "a research method validating self determination" make it a specialized book rather than a general personality text. This book can serve a wide audience mainly as a point of departure for much needed critical and unfettered discussion of psychology and human values. It raises questions about the relationship between normative and empirical approaches to the human person which all psychologists might well consider. The reviewer wishes it could have been possible for the writers to include a non-Catholic psychologist in their team to increase the objectivity of the presentation and to integrate their views even more with contemporary psychology. — *F. McK.*

Note. — The reviews were prepared by the Editor and the Associate Editors, who may be identified by their initials.

Burington, Richard Stevens, & May, Donald Curtis, Jr. *Handbook of probability and statistics with tables.* Sandusky, O.: Handbook Publishers, 1953. Pp. ix + 332. $4.50.

This handbook is prepared in the style of the handbooks which are so important to such fields as chemistry and engineering; the binding, format, and materials are most attractive. The text is well supplied with formulas and terse discussions of their meaning and application. The tables are quite comprehensive and include binomial distribution functions, Poisson distribution functions, Beta function ratios, distributions for the familiar statistics, trigonometric functions, natural and common logarithms, and many other functions. The style and discussions are distinctly technical and the mathematical level of the presentation probably exceeds the preparation of most psychologists. — *J. R. W.*

Fink, Harold Kenneth. *Mind and performance.* New York: Vantage Press, 1954. Pp. xi + 113. $3.00.

This is the report of an experimental study of learning ability in a number of animal forms. An Arrow Maze, consisting of four parallel pathways leading off from a neutral area, was employed for all forms. Four successive goals were used for each of the species, and the criterion for learning was 10 successful trials. Time limits were different for each species. Human subjects were blindfolded and the maze presented as a stylus maze. Results were treated to yield percentages of the learning ability of man, set at 100. The final rank order of the species tested was man, pig, dog, rat and chick, cat, turtles (4 species tested), with learning ability of the pig about half that of man. Pavlov's concept of experimental extinction is employed to explain behavior in learning successive goals. The author discusses the relationship of his results to Crile's suggestion that behavioral differences may be related to metabolic functions. — *A. R.*

Freud, Sigmund. *The origins of psychoanalysis: letters to Wilhelm Fliess, drafts and notes: 1887-1902.* New York. Basic Books, 1954. Pp. xi + 486. $6.75.

Freud's letters to Fliess covering the period from 1887 to 1902 constitute not only one of the most im-

portant publications in the field of psychology in years, but also a fascinating human document. Never intended for publication, the letters throw a light on Freud the person, the doctor, the scientist, and the philosopher. At one and the same time they show him as a human being and a genius, a boasting egotist and a disillusioned neurotic, a humanitarian, a devoted family man, and a tyrant with his colleagues. But most of all, they reveal the painful steps by which Freud developed his psychology — his keen observations, his tentative hypotheses, his probing, revising, accepting, and rejecting of concepts and theories until he was satisfied, and then frequently starting all over again because a new patient provided contradictions not previously met. Through the letters we follow him in his ups and downs, his satisfactions and disappointments, his self-analysis.

On October 16, 1895 he wrote, "I am practically sure I have solved the riddle of hysteria and obsessional neurosis"; on January 1, 1897, "You can see that I am in the full swing of discovery"; on May 31, 1897 he reports his first inkling of the Oedipus: "It seems as though in sons this death-wish is directed against their father and in daughters against their mother." And so it goes, discovery after discovery told in personal revelations: "I have jotted them down for your eyes alone." His first contact with infantile fantasy in neuroses, the development of libido, the mechanism of regression, the interpretation of dreams, the discovery of everyday psychopathology, of transference, and of many other psychological mechanisms are seen in Freud's first excitement at their unfolding. Included in the volume is Freud's "Project for a Scientific Psychology" which he wrote in 1895 and sent to Fliess for his opinion. This was an attempt to integrate neurology, psychology, and "objective" quantification into a "natural science." This project was later dropped by Freud, and many of its segments adapted in his numerous works. It is nevertheless historically important. A welcome by-product in this publication is Kris's masterful introduction, which throws much light on the letters. Kris's footnotes throughout relate the contents of the letters to Freud's better known works. — *M.K.*

Gill, Merton, Newman, Richard, Redlich, Fredrick C., & Sommers, Margaret. *The initial interview in psychiatric practice.* New York: International Universities Press, 1954. Pp. 423. $6.00. *Records,* 3, 12-in., 33 r.p.m., $4.60.

The first interview with a patient or client embarking on psychotherapy has puzzled experienced and novice therapists alike, and little has been written about it. This volume and its accompanying records present a viewpoint about the initial interview, and provide excellent materials for discussion and instruction. The authors are critical of the traditional psychiatric examination with its rigid structure and probing questions. They also do not entirely accept the position taken by nondirective therapists and some psychoanalysts that the opening interview involves no evaluation or diagnosis. Instead they attempt with considerable success to blend a diagnostic attitude — appraising the patient's need, motivation, and capacity for therapy — with a recognition that the first contact is an integral and often crucial part of the therapeutic process itself. The book consists of 119 pages of exposition and theory with a good review of the literature on the initial interview, and of 278 pages of verbatim illustrative material. Three initial interviews are given in full, with the transcripts on the left-hand pages and a parallel discussion on the right-hand pages. Two of the interviews, by experienced and able therapists, illustrate the application of the authors' approach. These interviews are not perfect, as the authors readily admit, but they are vivid and instructive. The remaining interview, conducted by an inexperienced student, shows most of the mistakes that can be made in interviewing — the interviewer's self-involvement, his inability to really hear the patient, and his defensive dependence on routine questions. The technically excellent phonograph records, available at cost to qualified professional persons and institutions, give the listener many cues that are lacking in the printed transcripts. The authors' moderate and eclectic conception of the initial interview will appeal to most readers as sound and widely applicable. The book and records supply unique and useful tools of instruction for clinical psychology and social work as well as for psychiatry. — *L. F. S.*

Humphreys, Anthony, & Traxler, Arthur E. *Guidance services.* Chicago: Science Research Associates, 1954. Pp. xvii + 438.

This is a text for an introductory course in guidance services. It covers the five subject matter areas reflected in many present courses with appropriately greater emphasis on basic understandings, tools and techniques, and on students' problems than on administration. The authors properly devote a chapter to the future of guidance services, and indicate the trend toward greater professionalization and increase in scientific emphasis. The swell of the trend seems to be definitely in the future if it includes reduction of the nomothetic "do and don't" presentation and an increase in a style which stimulates deeper understanding of the individual student. Humphreys and Traxler should be commended for devoting a quarter of their book to "solving students' major problems." The discussion does not, however, devote much space to students' cases. Student personnel and guidance services are both defined on the same page, but it is not entirely clear to the reader how these overlapping terms differ basically. This whole field of endeavor, as the authors intimate, is undergoing a rapid transition and the pioneer guidance workers in the field who have done much to improve the lot of individual students need to be better linked with the newly and more tech-

nically trained worker of the present. Books of this kind can do much to bring the two together. There is still, however, a great gap between the better books in guidance like this and the new volumes on counseling and personnel techniques. — *F. McK.*

Peck, Harris, B., & Bellsmith, Virginia. *Treatment of the delinquent adolescent.* New York: Family Service Association, 1954. Pp. 147 (paper). $2.00.

The authors have given us a realistic picture of the difficulties and successes in the use of psychotherapeutic techniques with delinquent adolescents. There is an excellent appraisal of the limitations and potentialities of group therapy, both with parents and with adolescents themselves. Educators, family and social agencies, and all those working with delinquents will find this book challenging reading. — *B.M.L.*

Warters, Jane. *Techniques of counseling.* New York: McGraw-Hill, 1954. Pp. viii + 384. $4.75.

This is a broad survey of the techniques available for student personnel work. It is written simply and well, and in a manner that will be helpful to the teacher or school counselor in training. The techniques fall largely under the headings of tests, observations, self-reports, status within the group, case study and conference, interview, and group methods. There is no attempt to survey the literature broadly, but the selection is quite appropriate. The reader is given samples of common techniques and an acquaintance with the uses, advantages, and disadvantages of each method. Moreover, a clear presentation of practical issues, especially those with ethical aspects, is included. There is some discussion of counseling, but the integration of several techniques into actual counseling situations is left for advanced clinical training. Miss Warters faces the dilemma of most writers in the field: how to encourage school personnel to use more objective techniques and at the same time instill sufficient caution concerning the limitations in the use of the techniques without supervised laboratory or clinical training. This problem is handled rather well. The reviewer, however, would have favored stronger and repeated emphasis on the necessity for more than verbal acquaintance with the techniques. — *F. McK.*

Wolberg, Lewis R. *The technique of psychotherapy.* New York: Grune & Stratton, 1954. Pp. xiv + 869. $14.75.

Good psychotherapy, says Dr. Wolberg in his preface, "must depend on a basically sound therapeutic structure, irrespective of the specified 'kind' of psychotherapy done and without reference to the individual 'style' of the therapist. The purpose of this book is to delineate such a structure." How well the book is judged to achieve this ambitious aim will depend on the experience and therapeutic predilections of those who read it. Few will deny, however, that this encyclopedic volume represents some of the most practical and down-to-earth writing about psychotherapy yet to appear in the literature. Rich in excerpts from recorded interviews and oriented around recurrent and realistic problems in therapeutic practice, it often has the flavor of a fruitful supervisory session. If its eclecticism and relative naiveté with respect to theory sometimes lead to contradictions and rather too easy solutions to difficult problems in the shadowland of therapeutic knowledge, the book nevertheless presents a large distillation of practical wisdom won from long and intensive experience. It makes possible an intimate acquaintance with one thoughtful and well-grounded point of view toward therapeutic conduct. One regrettable shortcoming, constituting a challenge to psychologists, is that in spite of its friendliness to psychology among the professions offering needed psychotherapeutic services, and its recognition of some of the contributions to therapeutic thought made by psychologists, this big book makes very little reference to the insightful investigative work done by psychologists or, indeed, to research in the field at all. As a summary of one therapist's considerable wisdom, *The Technique of Psychotherapy* is comprehensive and admirable. For secure knowledge, we shall not surprisingly have to wait. — *E. J. S.*

Wolstein, Benjamin. *Transference: its meaning and function.* New York: Grune & Stratton, 1954. Pp. xiii + 206. $5.00.

Written with sensitivity, scholarship, and logical sophistication, this careful historical analysis of the concept of transference represents a genuine contribution to psychoanalytic theory. In the tradition associated with Harry Stack Sullivan, the book develops the thesis that psychoanalysis is fundamentally the study of interpersonal processes, that the therapeutic utility of transference phenomena lies in their enabling both therapist and patient to arrive at greater insights into the latter's modes of relating to his fellows, and that the transference situation is not only a medium by which already established values may be re-examined but the seedbed of new values and new orientations toward human relationships. In the course of this development, Dr. Wolstein demonstrates that he has ideas, writes with style and a gentle persuasiveness, and competently levies on a quiet but impressive learning. It is a disappointment, however, to find a psychologist writing so authoritatively in this field with a total disregard for the bulk of pioneering research that has been the central contribution of his discipline to therapeutic knowledge. An integration of empirical findings with Dr. Wolstein's informed level of theorizing could be a vast boon to psychotherapeutic thought and practice. — *E. J. S.*

Woodworth, Robert S., & Schlosberg, Harold. *Experimental psychology.* (Rev. Ed.) New York: Holt, 1954. Pp. xi + 948. $8.95.

The classic *Experimental Psychology,* first published in 1938, has been given a thorough revision. Most chapters are rewritten, some are new, and all

are brought up to date. Over half of the 2,480 entries in the bibliography were not cited in the earlier edition. In addition to its values for general psychology, what kind of a foundation does this book provide for students who will specialize in clinical psychology? An appraisal from this limited viewpoint is generally favorable. The three chapters on emotion have been revised extensively, with level of activation as an organizing concept. Learning is treated so as to provide a good base for the applications of learning theory to personality development and psychotherapy. There is a new chapter on motivation. Dynamic aspects of perception and memory have been given increased recognition. Of more importance than these details, however, is the general atmosphere of the volume. Its consistently functional and objective attitude, its persuasive demonstrations of the values of experimental method, and its clear and interesting style of writing will lead students not only to learn experimental psychology but also to appreciate and enjoy it. — *L. F. S.*

Wulfeck, Joseph W., & Bennett, Edward M. *The language of dynamic psychology as related to motivation research.* New York: McGraw-Hill, 1954. Pp. 111. $4.00.

This volume is a glossary of terms commonly appearing in the field of motivation and apt to be meaningless or confusing to the lay reader. Each of the terms presented is accompanied by a definition furnished by Wulfeck and by some explanatory comment by Bennett. The volume is sponsored by the Advertising Research Foundation and is designed to "help the advertising man to keep his terminology and his understanding abreast of current activities in the field of motivation research." To the best of this reviewer's empathic judgment the book would seem to serve its purpose. The professional psychologist, however, will find little here to enlighten him. In fact, he may occasionally be jarred by expressions such as this, "Instinct (Psychiat): a wastebasket category for the origin of any behavior which is not learned." — *W. A. H.*

Books Received

Barrell, Joseph. *A philosophical study of the human mind.* New York: Philosophical Library, 1954. Pp. xii + 575. $6.00.

Beck, Samuel J. *The six schizophrenias.* Res. Monogr. No. 6. New York: American Orthopsychiatric Association, 1954. Pp. xiii + 238. $5.00.

Bischof, L. J. *Intelligence.* Doubleday Papers in Psychology, No. 5. Garden City, N. Y.: Doubleday, 1954. Pp. vi + 33 (paper). 85¢.

Braatøy, Trygve. *Fundamentals of psychoanalytic technique.* New York: Wiley, 1954. Pp. xi + 404. $6.00.

Campbell, Paul, & Howard, Peter. *Remaking men.* New York: Arrowhead Books, 1954. Pp. 126. $1.50.

Cronbach, Lee J. *Educational psychology.* New York: Harcourt, Brace, 1954. Pp. xxvii + 628. $7.50.

English, O. Spurgeon, & Finch, Stuart M. *Introduction to psychiatry.* New York: Norton, 1954. Pp. viii + 621. $7.00.

Frank, Lawrence K., Harrison, Ross, Hellersberg, Elizabeth, Machover, Karen, & Steiner, Meta. *Personality development in adolescent girls.* Monogr. Soc. Res. Child Develpm., 1951, 16, No. 53. New Orleans, La.: Child Development Publications, 1953. Pp. 316 (paper).

Frosch, John. (Ed.) *Annual survey of psychoanalysis.* Vol. II, 1951. New York: International Universities Press, 1954. Pp. xiv + 724. $10.00.

Fuller, John L. *Nature and nurture.* Doubleday Papers in Psychology, No. 4. Garden City, N. Y.: Doubleday, 1954. Pp. vii + 40 (paper). 85¢.

Haber, Heinz. *The physical environment of the flyer.* Maxwell AFB, Ala.: Air University, 1954. Pp. vi + 179.

Himes, Joseph S. *Social planning in America.* Doubleday Short Studies in Sociology. Garden City, N. Y.: Doubleday, 1954. Pp. ix + 59 (paper). 95¢.

Hodgson, Kenneth W. *The deaf and their problems.* New York: Philosophical Library, 1954. Pp. xx + 364. $6.00.

Ittelson, William H., & Cantril, Hadley. *Perception.* Doubleday Papers in Psychology, No. 7. Garden City, N. Y.: Doubleday, 1954. Pp. ix + 33 (paper). 85¢.

Joseph, Harry, & Zern, Gordon. *The emotional problems of children.* New York: Crown Publishers, 1954. Pp. ix + 310. $3.75.

Katz, Daniel, Cartwright, Dorwin, Eldersveld, Samuel, & Lee, Alfred M. (Eds.) *Public opinion and propaganda.* New York: Dryden Press, 1954. Pp. xx + 797. $6.25.

Lazarsfeld, Paul F. (Ed.) *Mathematical thinking in the social sciences.* Glencoe, Ill.: Free Press, 1954. Pp. 444. $10.00.

Leonard, Edith M., Van Deman, Dorothy D., & Miles, Lillian E. *Counseling with parents.* New York: Macmillan, 1954. Pp. vii + 330. $3.75.

Odlum, Doris M. *Psychology, the nurse and the patient.* (2nd Ed.) New York: Philosophical Library, 1954. Pp. 168. $4.75.

Pearson, Gerald H. J. *Psychoanalysis and the education of the child.* New York: Norton, 1954. Pp. x + 357. $5.00.

Piaget, Jean. *The construction of reality in the child.* New York: Basic Books, 1954. Pp. xiii + 386. $6.00.

Pickford, R. W. *The analysis of an obsessional.* New York: Norton: 1954. Pp. xii + 234. $4.00.

Sarason, Seymour B. *The clinical interaction.* New York: Harper, 1954. Pp. x + 425. $5.00.

Savage, Leonard J. *The foundations of statistics.* New York: Wiley, 1954. Pp. xv + 294. $6.00.

Schanck, Richard L. *The permanent revolution in science.* New York: Philosophical Library, 1954. Pp. xvi + 112. $3.00.

Senn, Milton J. E. (Ed.) *Problems of infancy and childhood, 1953.* New York: Josiah Macy Jr. Foundation, 1954. Pp. 196. $2.75.

Solomon, Joseph C. *A synthesis of human behavior*: New York: Grune & Stratton, 1954. Pp. xii + 265. $5.50.

Sullivan, Harry Stack. *The psychiatric interview.* New York: Norton, 1954. Pp. xxiii + 264. $4.50.

Taylor, W. S. *Dynamic and abnormal psychology.* New York: American Book Co., 1954. Pp. xiv + 658.

Watson, Robert I. *Psychology as a profession.* Doubleday Papers in Psychology, No. 8. Garden City, N. Y.: Doubleday, 1954. Pp. x + 65. 85¢.

NEUROLOGY & PSYCHIATRY

Section VIII of Excerpta Medica

The Neurology Section includes abstracts of all articles on neurological surgery in addition to general clinical neuro-physiology.

The section on Psychiatry covers the topics usually contained within the meaning of the word. In addition, the following special branches are dealt with fully:

* **Psychoanalysis, individual and analytical psychology**
* **Sexology, criminology, alcoholism, and drug addiction as related to psychiatry**
* **Social and industrial psychology and psychiatry, vocational guidance and mental testing**
* **Psychometrics**
* **Personality testing and the Rorschach test**
* **Heredity and statistical studies as they apply to psychiatry**
* **Mental defect and epilepsy**

Annual volume of 1000 pages, including an index classified by author and subject.

Subscription price $22.50 per annual volume

EXCERPTA MEDICA SERVICE CORPORATION

250 Madison Ave. **New York 16, N.Y.**

Please enter my subscription to Section VIII, NEUROLOGY AND PSYCHIATRY, OF EXCERPTA MEDICA. I enclose $22.50.

Name_____

Address_____

_____

Group Report of a Program of Research in Psychotherapy

Psychotherapy Research Group
The Pennsylvania State University

Nine integrated investigations of psychotherapy based on 43 recorded therapy cases and several control groups. In all, 218 recorded client-centered interviews, the work of 15 therapists, were analyzed. The clients received pre- and post-therapy tests and scales, developed specifically for this group of studies.

Part I

Chapter I. Description of the Psychotherapy Research Program at The Pennsylvania State University.

Part II

Chapter II. The Problem of Escaping Clients in Nondirective Counseling. *James J. Gallagher, University of Illinois*

Chapter III. A Study of the Relationships between Certain Counselor and Client Characteristics in Client-Centered Therapy. *Manuel Arvonson, Albany VA Hospital*

Chapter IV. Measuring Client Progress in Client-Centered Therapy. *John E. Tucker, VA Mental Hygiene Clinic, Albany, N. Y.*

Chapter V. The Role of Rorschach Variability in the Prediction of Client Behavior during Psychotherapy. *John M. Raimoto, West Haven VA Hospital.*

Chapter VI. The Role of Perceptual Consistency and Generality Change in Rorschach and Psychotherapy Behavior. *Marvin W. Kahn, Colorado University Medical School*

Chapter VII. An Assessment of the Predictive Value of Certain Language Measures in Psychotherapeutic Counseling. *Horace A. Page, University of Wisconsin*

Chapter VIII. The Type-Token Ratio as a Measure of Changes in Behavior Variability during Psychotherapy. *Jean J. G. Rosthal, University of Illinois*

Chapter IX. Verbal Signs of Resistance in Client-Centered Therapy. *John F. Gillespie, Jr., Roanoke VA Hospital*

Chapter X. A Comparison of More Improved with Less Improved Clients Treated by Client-Centered Methods. *Ben Ami Blau, Aberdeen Proving Grounds*

Part III

Chapter XI. General Summary and Discussion. *William U. Snyder, Pennsylvania State University*

Appendix I. Client Post-Therapy Rating Scale

Appendix II. Therapist Personality Rating Scale

Appendix III. Counselor Post-Therapy Rating Scale

Appendix IV. Counselor Coding Categories

Appendix V. Client Coding Categories

Appendix VI. Additional Scoring Notes for XCFu, XCFi, and XIT

Appendix VII. Verbal Signs of Resistance

Appendix VIII. Supplementary Correlations Relative to the Relationship between the Use of Client-Centered Techniques and the Success of Psychotherapy

Paper bound. 179 pages. Price $2.35, postpaid. Mail order and remittance to:
WILLIAM U. SNYDER, Department of Psychology, State College, Pennsylvania